PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07K 14/705, C12N 15/12, A61K 38/17, C12Q 1/68 — A1

(11) International Publication Number: WO 99/00422

(43) International Publication Date: 7 January 1999 (07.01.99)

(21) International Application Number: PCT/US98/13680

(22) International Filing Date: 30 June 1998 (30.06.98)

(30) Priority Data:
60/051,284 30 June 1997 (30.06.97) US

(71) Applicant: PRESIDENT AND FELLOWS OF HARVARD COLLEGE [US/US]; 124 Mt. Auburn Street, Cambridge, MA 02138 (US).

(72) Inventors: BUCK, Linda; Apartment 2, 27 Dartmouth Street, Boston, MA 02116 (US). DULAC, Catherine; 10 DeWolfe Street #48, Cambridge, MA 02138 (US). HERRADA, Gilles; 153 Salem Street, Boston, MA 02113 (US). MATSUNAMI, Hiroaki; 15 University Road #31, Brookline, MA 02146 (US).

(74) Agent: PLUMER, Elizabeth, R.; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
With amended claims.

(54) Title: NOVEL FAMILY OF PHEROMONE RECEPTORS

(57) Abstract

The invention describes a multigene family encoding a collection of novel mammalian pheromone receptors. Nucleic acids encoding the pheromone receptor polypeptides, including fragments and biologically functional variants thereof are provided. Also included are polypeptides and fragments thereof encoded by such nucleic acids, and antibodies relating thereto. Methods and products for using such nucleic acids and polypeptides also are provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

NOVEL FAMILY OF PHEROMONE RECEPTORS

Field of the Invention

This invention relates to nucleic acids and encoded polypeptides which are part of a multigene family encoding a collection of novel mammalian pheromone receptors. The invention further provides representative nucleic acids and encoded polypeptides in this multigene family. The representative polypeptides are expressed in the murine and rat vomeronasal organ (VNO). Agents which bind the nucleic acids or polypeptides also are provided. The invention further relates to methods of using such nucleic acids and polypeptides in the diagnosis and/or treatment of disease, including the use of these molecules in controlling fertility and behavior in vertebrates and invertebrates.

Background of the Invention

Pheromones are intraspecific chemical signals found throughout the animal kingdom. They regulate populations of animals by inducing innate behaviors and stereotyped changes in physiology (Karlson and Luscher, *Nature*, 1959, 183:55-56; Wilson, *Sci. Am.*, 1963, 208:100-114; Sorensen, *Chem. Sens.*, 1996, 21:245-256). Pheromones can serve as cues for overcrowding, impending danger, reproductive status, gender, or dominance. In rodents, a variety of pheromone effects have been reported. These include effects on estrus and the onset of puberty as well as the induction of mating and aggressive behaviors (Singer, A.G., *J. Steroid. Biochem. Molec. Biol.*, 1991, 39:627-632; Halpern, M., *Ann. Rev. Neurosci.*, 1987 10:325-362; Wysocki, C.J., et al., *In the Neurobiology of Taste and Smell*, 1987, 125-150; Novotny et al., *Chemical signals in Vertebrates*, 1990, Vol. 5, eds. D.W. Macdonald et al., Oxford University Press).

The detection of pheromones is mediated by the olfactory system. However, sensory neurons that detect pheromones are typically segregated from those that detect volatile odorants (Keverne, E.B., *Trends Neurosci.*, 1983, 6:381-384; Halpern, M., *Ann. Rev. Neurosci.*, 1987, 10:325-362; Wysocki, C.J., et al., *In the Neurobiology of Taste and Smell*, 1987, 125-150; Hildebrand, J.G., et al., *Brain Res.*, 1997, 677:157-161). In mammals, sensory neurons in the nasal olfactory epithelium (OE) detect volatile odorants and some pheromones while those in an

accessory olfactory organ, called the vomeronasal organ (VNO), are thought to be specialized to detect pheromones. The VNO is a tubular structure, at the base of the nasal septum, which is connected to the nasal cavity by a small duct. Signals from the OE are relayed through the olfactory bulb (OB) to the olfactory cortex, and then to multiple brain regions, including those involved in conscious perception. In contrast, signals from the VNO are conveyed through the accessory olfactory bulb (AOB) to the amygdala and hypothalamus, areas associated with the endocrine and behavioral responses induced by pheromones.

Volatile odorants are detected in the OE by as many as 1000 different types of odorant receptors (ORs), which are differentially expressed by olfactory sensory neurons (Buck and Axel, *Cell*, 1991, 65:175-187; Levy, N.S., et al., *J. Steroid Biochem. Mol. Biol.*, 1991, 39:633-637, 1991; Nef, P., et al., *Proc. Natl. Acad. Sci.*, 1992, 89:8948-8952; Strotman, J., et al., *Neuroreport*, 1992, 3:1053-1056; Ngai, J., et al., *Cell*, 1993, 72:667-680; Ressler, K.J., et al, *Cell*, 1993, 73:597-609; Vassar, R., et al, *Cell*, 1993, 74:309-318. The ORs are thought to couple to the G protein a subunit, $G\alpha_{olf}$, thereby initiating a cascade of transduction events which culminate in the generation of action potentials in the sensory axons (reviewed in Firestein, S., *Curr.Opin. in Neurobiology*, 1992, 2:444-448; Reed, R., *Neuron*, 1992, 8:205-209; Ronnett, G., et al., *Trends Neurosci*, 1992, 15:508-513). Current evidence suggests that each OR may recognize a particular molecular feature that can be shared by many odorants (Ressler, K., et al., *Cell*, 1994, 79:1245-1255; Vassar, R., et al., *Cell*, 1994, 79:981-991; Axel, R., *Sci. Am.*, 1995, 1273:154-159; Buck, L., *Annu. Rev. Neurosci.*, 1996, 19:517-544). This is consistent with a combinatorial coding model in which the identities of different odorants are encoded by different combinations of receptors, but each receptor serves as one component of the codes for many odorants. By contrast, very little is known about how pheromones are detected or encoded in the VNO. Although VNO neurons (VNs) resemble olfactory sensory neurons in the nose, only a rare VN expresses an OR gene. VNs also lack a number of other olfactory sensory transduction molecules, including the G protein a subunit,$G\alpha_{olf}$ (Reed, R., *Neuron*, 1992, 8:205-209), which is highly expressed in olfactory neurons (Dulac and Axel, *Cell*, 1995, 83:195-206; Berghard, A., et al, *Proc. Natl. Acad. Sci. USA*, 1996, 93:2365-2369; Wu, Y., et al, *Biochem. Biopys. Res. Com.*, 1996, 220:900-904). Instead, VNs express high levels of two other G protein a subunits,$G\alpha_0$ and $G\alpha i_2$ (Dulac and Axel, *Cell*, 1995, 83:195-206; Halpern, M., *Brain Res.*, 41995, 677:157-161; Berghard, A., et al, *Proc. Natl. Acad. Sci. USA*, 1996, 93:2365-2369). G_{α_0} and $G\alpha i_2$ are expressed in spatially-segregated subsets of VNs that form longitudinal zones

in the VNO neuroepithelium. Interestingly, Dulac and Axel have identified a family of ~100 candidate pheromones receptors ("VNRs") which appear to be expressed exclusively in the $G\alpha i_2$ subset (Dulac and Axel, *Cell*, 1995, 83:195-206).

This invention differs from the state of the art in providing a novel family of mammalian pheromone receptors. Accordingly, the objects of the invention relate to providing compositions containing these novel receptors and their binding partners and methods for using such compositions to modulate pheromone receptor activity.

Summary of the Invention

The invention involves the discovery of a multigene family of mammalian pheromone receptors. In particular, the invention involves the cDNA cloning of multiple pheromone receptors from a murine VNO cDNA library and from a rat VNO cDNA library. Partial sequences of human homologs of these pheromone receptors also are provided.

In general, the invention provides isolated nucleic acid molecules encoding the novel pheromone receptors, unique fragments of the isolated nucleic acid molecules, expression vectors containing the foregoing, and host cells transfected with the foregoing. The invention also provides isolated pheromone receptor polypeptides and agents which bind such polypeptides, including antibodies. The foregoing can be used in the diagnosis or treatment of conditions, including the control of fertility, that are characterized by the expression of a pheromone receptor polypeptide. Methods for identifying pharmacological agents useful in the diagnosis or treatment of such conditions and methods for identifying additional members of this multigene family also are provided.

Applicants have discovered that the pheromone receptors disclosed herein are expressed in the vomeronasal organ (VNO), particularly in $G\alpha_0$ protein expressing neurons. This is in contrast to the prior art VNO pheromone receptors which are expressed in neurons which express different G-coupled proteins ($G\alpha i_2$-expressing neurons). Thus, the novel pheromone receptors disclosed herein are distinct from, and expressly exclude, the prior art VNO pheromone receptors which differ in primary structure, as well as in cell localization. Although Applicants do not intend the invention to be limited to a particular theory or mechanism, the amino acid sequence homology and structural organization of the pheromone receptor polypeptides to other well-known G-protein coupled receptors suggests that the pheromone receptors disclosed herein also are G-protein coupled. Thus, it is anticipated that the binding to the pheromone receptor of its

cognate ligand (pheromone) will be accompanied by G-protein signal transduction, an event which can be measured using conventional screening assays, such as assays that measure changes in the intracellular concentrations of calcium and/or cyclic nucleotides (see, e.g., PCT publication no. WO 94/18959, entitled "Calcium Receptor-Active Molecules", inventors E. Nemeth et al.).

According to one aspect of the invention, a family of pheromone receptor polypeptides is provided. Each polypeptide of the family shares amino acid sequence homology and structural organization with a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52. Each polypeptide member of the receptor family contains, from amino terminus to carboxyl terminus, the following domains: (a) an amino-terminal extracellular domain containing from 30 to 600 amino acids; (b) a transmembrane region comprising: (i) seven non-contiguous transmembrane domains designated TM1, TM2, TM3, TM4, TM5, TM6 and TM7, (ii) three non-contiguous extracellular domains designated EC2, EC3 and EC4, and (iii) three non-contiguous intracellular domains designated IC1, IC2, and IC3,wherein the transmembrane domains, the extracellular domains and the intracellular domains are attached to one another from amino terminus to carboxyl terminus in the order TM1-IC1-TM2-EC2-TM3- IC2-TM4-EC3-TM5-IC3-TM6-EC4-TM7, and wherein the transmembrane region has at least about 35% homology and a length approximately equal to a transmembrane region of a polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 34, 36, 38, 40, 42, 44, 46, 48, and 50; and (c) a carboxyl-terminal intracellular domain containing from 5 to 200 amino acids. Each polypeptide member of the family is expressed in a $G\alpha_o$ protein-expressing vomeronasal organ neuron or are expressed in another olfactory organ neuron in an animal which does not possess a vomeronasal organ. One skilled in the art can readily identify olfactory organs in animals which do not possess a vomeronasal organ.

In general, the amino-terminal extracellular domains (NTDs) of the receptor family members share sequence homology to a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, and 50 to a lesser extent than that observed for the transmembrane region. The length of the extracellular domain can vary among members of the family. Accordingly, certain embodiments of the invention have extracellular domains that contain at least 50, 100, 200, 300, 400 or 500 amino acids. Preferably, the transmembrane region has greater than 40% homology

with the corresponding region of a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 34, 36, 38, 40, 42, 44, 46, 48, and 50, and more preferably, have even greater sequence homology (e.g., more than 50%, 60%, 70%, 80% or 90% homology). The length of the carboxyl-terminal intracellular domain can vary among members of the family. Accordingly, certain embodiments of the invention have carboxyl-terminal intracellular domains that contain at least between 5 and 50 amino acids. More preferably, carboxyl-terminal intracellular domains contain between 15 and 25 amino acids.

According to another aspect of the invention, a method for identifying a nucleic acid encoding a pheromone receptor is provided. The method involves contacting a mixture of nucleic acid molecules (genomic library, cDNA library, genomic DNA, RNA, etc.) with at least one nucleic acid probe of a nucleic acid selected from the group consisting of: (a) a nucleic acid molecule selected from the group consisting of SEQ ID NO. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 54, and 55 that encodes a pheromone receptor polypeptide; (b) a unique fragment of (a); (c) a human homolog of (a) or (b); and (d) a set of degenerate primers of any of (a), (b) or (c); and identifying the sequences within the mixture that hybridize to the probe. Selected fragments of human homologs of a pheromone receptor are selected from the group consisting of SEQ ID NO. 51, 53, 54 and 55. In certain embodiments, the nucleic acid probe further includes a detectable label to facilitate identification of the sequence in the library which hybridizes to the probe. In certain embodiments, the probe is represented by a pair of degenerate polymerase chain reaction ("PCR") primers that amplify a unique fragment of a nucleic acid molecule selected from the group consisting of SEQ ID NO. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 54, and 55. The meaning of "unique fragment" in reference to a nucleic acid is provided below. By "degenerate PCR primers that amplify a unique fragment" is meant degenerate primers which result in the amplification of a unique fragment following a polymerase chain reaction. According to this embodiment, the method for identifying a nucleic acid encoding a pheromone receptor polypeptide further involves subjecting a mixture of nucleic acids and the degenerate PCR primers to amplification conditions prior to identifying the sequences of the mixture that hybridize to the probe and that form part of the amplification reaction products. In some embodiments the pair of degenerate polymerase chain reaction primers is selected from a conserved sequence motif of a pheromone receptor polypeptide. A "conserved sequence motif" can be determined using the side-by-side comparison of the amino acid sequences of the different

pheromone receptor polypeptides of the invention. Exemplary conserved sequence motifs include regions selected from the group consisting of amino acids 191-397, amino acids 565-825, amino acids 637-825, amino acids 637-804, amino acids 619-784, of the polypeptide of, for example, SEQ ID NO. 2 (VR1). In preferred embodiments, the pair of degenerate polymerase chain reaction primers is selected from the group consisting of SEQ ID NOs. 60 and 61, SEQ ID NOs. 62 and 63, SEQ ID NOs. 64 and 63, SEQ ID NOs. 64 and 65, and SEQ ID NOs. 66 and 67.

According to yet another aspect of the invention, an isolated nucleic acid molecule is provided. The isolated nucleic acid molecule hybridizes under high or low stringency conditions to a molecule consisting of a nucleic acid sequence selected from the group consisting of SEQ ID NO. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 54, and 55. The invention further embraces nucleic acid molecules that differ from the foregoing isolated nucleic acid molecules in codon sequence due to the degeneracy of the genetic code. The invention also embraces complements of the foregoing nucleic acids.

The pheromone receptors of the invention are expressed in the vomeronasal organ or, in an animal which lacks such an organ, are expressed in another olfactory organ. More particularly, the receptors of the invention are expressed in a $G\alpha_o$ protein-expressing vomeronasal organ neuron. Although not intending to be bound to a particular mechanism, it is believed that the receptors of the invention are G-protein coupled receptors. This is supported by Applicants' discovery that the receptors of the invention are expressed in $G\alpha_o$ protein-expressing vomeronasal organ neurons.

The pheromone receptors of the invention bind to ligands (pheromones) which induce certain changes in receptor conformation. Methods for identifying ligands which bind to the pheromone receptors of the invention are provided below, e.g., by forming an affinity matrix containing immobilized receptor and using the matrix to isolate a cognate ligand from a complex mixture. The particular ligand bound by a particular receptor is dictated by the primary and secondary structure of the receptor. In certain embodiments, the immobilized pheromone receptor polypeptide is a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52.

According to another aspect of the invention, an isolated nucleic acid molecule that is a unique fragment of any of the foregoing isolated nucleic acid molecules is provided. In general, the isolated nucleic acid molecule consists of a unique fragment between 12 and 4000

nucleotides in length, and complements thereof, of any cDNA (SEQ ID NOs. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 54, and 55) encoding a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52. Depending upon its intended use (e.g., probe, primer), the unique fragment can be between 12 and 2000, 1000, 500, 250, 100, 50 or 25 nucleotides in length. Preferably, the isolated nucleic acid molecule consists of between 12 and 35 contiguous nucleotides of the foregoing cDNAs encoding the pheromone receptor polypeptides, or complements of such nucleic acid molecules. More preferably, the unique fragment is at least 14, 15, 16, 17, 18, 20 or 22 contiguous nucleotides of the nucleic acid sequence of the foregoing cDNAs encoding the pheromone receptor polypeptides, or complements thereof. Particularly preferred isolated nucleic acid molecules are isolated fragments of the foregoing cDNAs which encode one or more of the following pheromone receptor polypeptide domains, alone or in combination (e.g., as fusion proteins): an amino-terminal extracellular domain, a transmembrane region, and a carboxy-terminal intracellular domain. In certain embodiments, the unique fragments are a pheromone receptor extracellular domain or a pheromone receptor intracellular domain coupled to at least one (e.g., 1, 2, 3, 4, 5, 6, or 7) transmembrane domain.

According to yet another aspect of the invention, an isolated nucleic acid molecule comprising a molecule having a sequence selected from the group consisting of SEQ ID NO. 51, 53, 54, 55, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, and 92, that encodes a pheromone receptor polypeptide are provided. This aspect of the invention further embraces nucleic acid molecules that differ from these nucleic acid molecules in codon sequence due to the degeneracy of the genetic code, and diversity among pheromone receptors and complements of foregoing.

According to still other aspects of the invention, an expression vector comprising any of the foregoing isolated nucleic acid molecules operably linked to a promoter and host cells transformed or transfected with the same also are provided.

According to another aspect of the invention, an isolated polypeptide encoded by any of the above-described isolated nucleic acid molecules is provided. Preferably, the isolated polypeptide is a pheromone receptor polypeptide that has a pheromone receptor activity or an antigenic fragment thereof. As used herein, a pheromone receptor activity refers to the ability of the pheromone receptor to selectively bind to its cognate ligand (pheromone) and, optionally,

upon binding, to induce signal transduction in a cell that expresses the pheromone receptor. In preferred embodiments, the isolated polypeptide comprises a pheromone receptor polypeptide having a sequence selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52.

According to yet other embodiments, the isolated polypeptide comprises a polypeptide encoded by a nucleic acid which hybridizes under high or low stringency conditions to the extracellular domain, transmembrane region and/or intracellular domain of a cDNA sequence selected from the group consisting of SEQ ID NO. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 54, and 55 that encodes a pheromone receptor polypeptide or fragment thereof. Thus, the invention embraces portions of a pheromone receptor polypeptide that may include, for example, an amino-terminal extracellular domain or a carboxy-terminal intracellular domain coupled to 1, 2, 3, 4, 5, 6, or 7 transmembrane domains. Preferably, such polypeptides or fragments thereof are unique fragments and can function as, for example, antigens for making antibodies specific for pheromone receptor family members. Accordingly, the polypeptides of the invention can be used to isolate additional members of the pheromone receptor family or, alternatively, can be used to induce in vivo an immune response to a pheromone receptor, i.e., can be incorporated into a vaccine preparation. Such vaccine compositions are useful for controlling fertility or behavior in an animal by administering to the animal, an effective amount of the vaccine to elicit an immune response to the pheromone receptor. Thus, the invention embraces fragments or variants of the foregoing pheromone receptors which exhibit certain detectable activities, e.g., a ligand binding activity, an antigenicity activity. In certain embodiments, the isolated polypeptide is encoded by a cDNA selected from the group consisting of SEQ ID NO. 51, 53, 54, 55, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, and 92, that encodes a pheromone receptor polypeptide or one or more of its domains.

According to another aspect of the invention, there are provided isolated binding polypeptides which selectively bind a unique amino acid sequence of a pheromone receptor polypeptide or fragment thereof. The isolated binding polypeptide in certain embodiments binds to a polypeptide comprising the extracellular domain and/or 1, 2, 3, 4, 5, 6, or 7 transmembrane domains of a pheromone receptor polypeptide sel cted from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52.

The isolated polypeptide preferably binds to a polypeptide consisting of the amino-terminal extracellular domain and/or one or more portions of the transmembrane region of a pheromone receptor polypeptide sequence selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52. In preferred embodiments, isolated binding polypeptides include antibodies and fragments of antibodies (e.g., Fab, $F(ab)_2$, Fd and antibody fragments which include a CDR3 region which binds selectively to the unique sequences of the polypeptides of the invention). In the preferred embodiments, the isolated binding peptides do not bind to pheromone receptors that are expressed in vomeronasal organ neurons other than Gαo-protein-expressing neurons.

The invention provides in yet other aspects, isolated nucleic acids or polypeptides of the invention that are: (a) immobilized to an insoluble support (an affinity matrix containing immobilized pheromone receptor polypeptide or a unique fragment thereof); (b) associated with, covalently coupled to, or encapsulated a drug delivery device (e.g., a microsphere) to effect controlled release of the isolated nucleic acid or polypeptide in vivo or in vitro; (c) covalently coupled to another isolated nucleic acid or protein to form a chimeric molecule; and/or (d) labeled with a detectable agent (e.g., a radiolabel, a fluorescent label). Thus, the invention provides chimeric molecules containing at least one first structural domain of one pheromone receptor polypeptide (e.g., an extracellular domain) coupled to a second structural domain (e.g., a transmembrane domain, such as TM1, TM2, etc.) of a different pheromone receptor polypeptide. The invention also provides a method for isolating a pheromone receptor by (1) contacting a composition containing a putative pheromone receptor of the above-described family with an affinity matrix containing immobilized binding polypeptide under conditions to permit the pheromone receptor to selectively bind to the immobilized binding polypeptide, and (2) isolating the polypeptides that bind to the affinity matrix.

According to still another aspect of the invention, pharmaceutical compositions containing any of the foregoing compounds of the invention in a pharmaceutically acceptable carrier and methods of producing same by placing the compositions in the carrier also are provided.

According to still another aspect of the invention, methods for modulating a pheromone receptor activity (e.g., a ligand binding activity, a signal transduction activity) in a cell (vertebrate or invertebrate) are provided. The cell can be located in vivo or in vitro and the methods can be used to down regulate (inhibit) or up regulate (stimulate) the pheromone receptor

activity. For example, to inhibit a ligand binding activity, the cell is contacted with an inhibitor that can be an isolated binding polypeptide that binds to an extracellular portion of the receptor and, thereby, inhibits receptor binding to its cognate ligand. Such binding also can induce conformational changes in the receptor that alter the signal transduction activity of the receptor. The inhibitor can be an isolated antibody (or function equivalent thereof) which binds to an epitope located on an extracellular portion (such as EC2, EC3, EC4) of the pheromone receptor polypeptide, e.g., an amino-terminal extracellular domain or an "extracellular transmembrane region domain", i.e., an extracellular portion of the transmembrane region located between one or more transmembrane domains. Alternatively, the inhibitor can be an agent (e.g., an isolated competitive binding polypeptide) that inhibits receptor-ligand binding. For example, the inhibitor can be an isolated fragment of a pheromone receptor (preferably, a soluble fragment), which fragment contains a ligand (pheromone) binding site. Other inhibitors can be identified in screening assays which test the ability of a putative inhibitor to inhibit pheromone receptor-mediated signal transduction or which test the ability of the putative inhibitor to inhibit binding of a pheromone receptor to its known cognate ligand. Similarly, such screening assays can be used to identify molecules which stimulate pheromone receptor-mediated signal transduction. Exemplary molecules which stimulate transduction include the naturally-occurring ligands (e.g., isolated from a biological source (e.g., urine, vaginal fluid), as well as synthetic ligands obtained from a non-biological source (e.g., a combinatorial library).

According to still another aspect of the invention, methods for inhibiting the binding of a pheromone having a binding domain to a pheromone receptor polypeptide having a ligand binding site that selectively binds to the binding domain are provided. The method involves contacting (in vivo or in vitro) the pheromone receptor polypeptide with an agent which binds to the ligand binding site under conditions to permit binding of the agent to the receptor. For example, the agent can be an isolated binding polypeptide that binds to the ligand binding site of the pheromone receptor. Thus, the agent can be an isolated antibody (or functionally equivalent fragment thereof) which selectively binds to the ligand binding site of the receptor. Alternatively, the agent can be a pheromone receptor antagonist, e.g., a molecule that mimics the structure of the naturally-occurring ligand but that does not mimic the function (stimulating the receptor) of the naturally-occurring ligand. Agents which inhibit ligand binding can be identified in screening assays which test the ability of a putative binding inhibitor to inhibit

binding of a pheromone receptor to its cognate ligand (e.g., pheromone). Such molecules can be isolated from a biological source or from a non-biological source.

According to another aspect of the invention, methods for modulating pheromone receptor-mediated signal transduction in a subject are provided. The methods involve administering to a subject in need of such treatment an agent that selectively binds to any of the above-described isolated nucleic acid molecules which encode a pheromone receptor or unique fragment thereof, or an expression product thereof, in an amount effective to modulate (down regulate or up regulate) pheromone receptor-mediated signal transduction in the subject. Exemplary agents include antisense nucleic acid molecules and binding polypeptides.

Thus, according to yet another aspect of the invention, methods are provided for identifying lead compounds for an pharmacological agent useful in the diagnosis or treatment of a condition associated with pheromone receptor signal transduction activity or otherwise generally associated with binding of the receptor to its cognate ligand. Preferably, cells expressing intact pheromone receptor polypeptides or portions thereof are used in the screening assays for identifying lead compounds which modulate pheromone receptor-mediated ligand binding or signal transduction activity. Cells expressing these polypeptides, isolated pheromone receptor polypeptides and fragments of these polypeptides which contain the ligand binding site can be used in the screening assays for identifying lead compounds which modulate binding of the receptor to a known ligand.

The screening methods involve forming a mixture of a pheromone receptor polypeptide (as noted above) or fragment thereof containing a ligand binding site; a molecule which is known to (1) interact with the foregoing receptor to effect pheromone receptor-mediated signal transduction or (2) bind to the ligand binding site of the receptor; and a candidate pharmacological agent. The mixture is incubated under conditions which, in the absence of the candidate pharmacological agent, permit a first amount of pheromone receptor-ligand binding or receptor-mediated signal transduction by the known ligand. A test amount of the selective binding of the ligand by receptor or of the specific activation of signal transduction is determined. Detection of an increase in the foregoing activities in the presence of the candidate pharmacological agent indicates that the candidate pharmacological agent is a lead compound for a pharmacological agent which increases specific activation of pheromone receptor-mediated signal transduction or selective binding of the ligand by the ligand binding site of the receptor. Detection of a decrease in the foregoing activities in the presence of the candidate

pharmacological agent indicates that the candidate pharmacological agent is a lead compound for a pharmacological agent which decreases specific activation of pheromone receptor-mediated signal transduction or selective binding of the ligand by the ligand binding site of the receptor.

Pheromone receptor polypeptides that are useful in the screening assays, preferably, are those selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52. Extracellular domains or portions thereof and portions of the transmembrane region, alone or coupled to one another, of these pheromone receptor polypeptides (indicated in the Examples) can be tested for their ability to inhibit receptor-ligand binding.

These and other objects of the invention will be described in further detail in connection with the detailed description of the invention.

All patents, patent publications, references and other information identified in this document are incorporated in their entirety herein by reference.

Brief Description of the Drawings

Figure 1 depicts a comparison of the deduced protein sequences encoded by VR cDNA clones.

Figure 2 is a schematic comparison of ORs, VNRs, and Vrs.

Figure 3 depicts a comparison of the deduced protein sequences encoded by the Go-VN cDNA clones.

Brief Description of the Sequences

SEQ ID NO. 1 is the nucleotide sequence of the mouse pheromone receptor VR1 cDNA (GenBank Accession No. AF011411).

SEQ ID NO. 2 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR1 cDNA (GenBank Accession No. AF011411).

SEQ ID NO. 3 is the nucleotide sequence of the mouse pheromone receptor VR2 cDNA (GenBank Accession No. AF011412).

SEQ ID NO. 4 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR2 cDNA (GenBank Accession No. AF011412).

SEQ ID NO. 5 is the nucleotide sequence of the mouse pheromone receptor VR3 cDNA (GenBank Accession No. AF011413).

SEQ ID NO. 6 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR3 cDNA (GenBank Accession No. AF011413).

SEQ ID NO. 7 is the nucleotide sequence of the mouse pheromone receptor VR4 cDNA (GenBank Accession No. AF011414).

SEQ ID NO. 8 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR4 cDNA (GenBank Accession No. AF011414).

SEQ ID NO. 9 is the nucleotide sequence of the mouse pheromone receptor VR5 cDNA (GenBank Accession No. AF011415).

SEQ ID NO. 10 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR5 cDNA (GenBank Accession No. AF011415).

SEQ ID NO. 11 is the nucleotide sequence of the mouse pheromone receptor VR6 cDNA (GenBank Accession No. AF011416).

SEQ ID NO. 12 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR6 cDNA (GenBank Accession No. AF011416).

SEQ ID NO. 13 is the nucleotide sequence of the mouse pheromone receptor VR7 cDNA (GenBank Accession No. AF011417).

SEQ ID NO. 14 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR7 cDNA (GenBank Accession No. AF011417).

SEQ ID NO. 15 is the nucleotide sequence of the mouse pheromone receptor VR8 cDNA (GenBank Accession No. AF011418).

SEQ ID NO. 16 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR8 cDNA (GenBank Accession No. AF011418).

SEQ ID NO. 17 is the nucleotide sequence of the mouse pheromone receptor VR9 cDNA (GenBank Accession No. AF011419).

SEQ ID NO. 18 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR9 cDNA (GenBank Accession No. AF011419).

SEQ ID NO. 19 is the nucleotide sequence of the mouse pheromone receptor VR10 cDNA (GenBank Accession No. AF011420).

SEQ ID NO. 20 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR10 cDNA (GenBank Accession No. AF011420).

SEQ ID NO. 21 is the nucleotide sequence of the mouse pheromone receptor VR11 cDNA (GenBank Accession No. AF011421).

SEQ ID NO. 22 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR11 cDNA (GenBank Accession No. AF011421).

SEQ ID NO. 23 is the nucleotide sequence of the mouse pheromone receptor VR12 cDNA (GenBank Accession No. AF011422).

SEQ ID NO. 24 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR12 cDNA (GenBank Accession No. AF011422).

SEQ ID NO. 25 is the nucleotide sequence of the mouse pheromone receptor VR13 cDNA (GenBank Accession No. AF011423).

SEQ ID NO. 26 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR13 cDNA (GenBank Accession No. AF011423).

SEQ ID NO. 27 is the nucleotide sequence of the mouse pheromone receptor VR14 cDNA (GenBank Accession No. AF011424).

SEQ ID NO. 28 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR14 cDNA (GenBank Accession No. AF011424).

SEQ ID NO. 29 is the nucleotide sequence of the mouse pheromone receptor VR15 cDNA (GenBank Accession No. AF011425).

SEQ ID NO. 30 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR15 cDNA (GenBank Accession No. AF011425).

SEQ ID NO. 31 is the nucleotide sequence of the mouse pheromone receptor VR16 cDNA (GenBank Accession No. AF011426).

SEQ ID NO. 32 is the predicted amino acid sequence of the polypeptide encoded by the mouse pheromone receptor VR16 cDNA (GenBank Accession No. AF011426).

SEQ ID NO. 33 is the nucleotide sequence of the rat pheromone receptor Go-VN1 cDNA (GenBank Accession No. AF016178).

SEQ ID NO. 34 is the predicted amino acid sequence of the polypeptide encoded by the rat pheromone receptor Go-VN1 cDNA (GenBank Accession No. AF016178).

SEQ ID NO. 35 is the nucleotide sequence of the rat pheromone receptor Go-VN2 cDNA (GenBank Accession No. AF016179).

SEQ ID NO. 36 is the predicted amino acid sequence of the polypeptide encoded by the rat pheromone receptor Go-VN2 cDNA (GenBank Accession No. AF016179).

SEQ ID NO. 37 is the nucleotide sequence of the rat pheromone receptor Go-VN3 cDNA (GenBank Accession No. AF016180).

SEQ ID NO. 38 is the predicted amino acid sequence of the polypeptide encoded by the rat pheromone receptor Go-VN3 cDNA (GenBank Accession No. AF016180).

SEQ ID NO. 39 is the nucleotide sequence of the rat pheromone receptor Go-VN4 cDNA (GenBank Accession No. AF016181).

SEQ ID NO. 40 is the predicted amino acid sequence of the polypeptide encoded by the rat pheromone receptor Go-VN4 cDNA (GenBank Accession No. AF016181).

SEQ ID NO. 41 is the nucleotide sequence of the rat pheromone receptor Go-VN5 cDNA (GenBank Accession No. AF016182).

SEQ ID NO. 42 is the predicted amino acid sequence of the polypeptide encoded by the rat pheromone receptor Go-VN5 cDNA (GenBank Accession No. AF016182).

SEQ ID NO. 43 is the nucleotide sequence of the rat pheromone receptor Go-VN6 cDNA (GenBank Accession No. AF016183).

SEQ ID NO. 44 is the predicted amino acid sequence of the polypeptide encoded by the rat pheromone receptor Go-VN6 cDNA (GenBank Accession No. AF016183).

SEQ ID NO. 45 is the nucleotide sequence of the rat pheromone receptor Go-VN7 cDNA (GenBank Accession No. AF016184).

SEQ ID NO. 46 is the predicted amino acid sequence of the polypeptide encoded by the rat pheromone receptor Go-VN7 cDNA (GenBank Accession No. AF016184).

SEQ ID NO. 47 is the nucleotide sequence of the rat pheromone receptor Go-VN13C cDNA (GenBank Accession No. AF016185).

SEQ ID NO. 48 is the predicted amino acid sequence of the polypeptide encoded by the rat pheromone receptor Go-VN13C cDNA (GenBank Accession No. AF016185).

SEQ ID NO. 49 is the nucleotide sequence of the rat pheromone receptor Go-VN13B cDNA (GenBank Accession No. AF016186).

SEQ ID NO. 50 is the predicted amino acid sequence of the polypeptide encoded by the rat pheromone receptor Go-VN13B cDNA (GenBank Accession No. AF016186).

SEQ ID NO. 51 is a partial nucleotide sequence of the human pheromone receptor hVR1.

SEQ ID NO. 52 is the predicted amino acid sequence of the polypeptide encoded by the partial sequence of the human pheromone receptor hVR1.

SEQ ID NO. 53 is a partial nucleotide sequence of the human pheromone receptor hVNO1.

SEQ ID NO. 54 is a partial nucleotide sequence of the human pheromone receptor hVNO2.

SEQ ID NO. 55 is a partial nucleotide sequence of the human pheromone receptor hVNO3.

SEQ ID NO. 56 is the nucleotide sequence of primer AL1.

SEQ ID NO. 57 is the nucleotide sequence of primer AL3.

SEQ ID NO. 58 is a fifty amino acid sequence of Go-VN13B (SEQ ID NO. 50) that is absent from Go-VN13C (SEQ ID NO. 48).

SEQ ID NO. 59 is the amino acid sequence of a rat kidney extracellular calcium/ polyvalent cation-sensing receptor.

SEQ ID NO. 60 is a degenerate oligonucleotide primer from a conserved VR domain.

SEQ ID NO. 61 is a degenerate oligonucleotide primer from a conserved VR domain.

SEQ ID NO. 62 is a degenerate oligonucleotide primer from a conserved VR domain.

SEQ ID NO. 63 is a degenerate oligonucleotide primer from a conserved VR domain.

SEQ ID NO. 64 is a degenerate oligonucleotide primer from a conserved VR domain.

SEQ ID NO. 65 is a degenerate oligonucleotide primer from a conserved VR domain.

SEQ ID NO. 66 is a degenerate oligonucleotide primer from a conserved VR domain.

SEQ ID NO. 67 is a degenerate oligonucleotide primer from a conserved VR domain.

SEQ ID NO. 68 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR1.

SEQ ID NO. 69 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR2.

SEQ ID NO. 70 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR3.

SEQ ID NO. 71 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR4.

SEQ ID NO. 72 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR5.

SEQ ID NO. 73 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR6.

SEQ ID NO. 74 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR7.

SEQ ID NO. 75 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR8.

SEQ ID NO. 76 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR9.

SEQ ID NO. 77 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR10.

SEQ ID NO. 78 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR11.

SEQ ID NO. 79 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR12.

SEQ ID NO. 80 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR13.

SEQ ID NO. 81 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR14.

SEQ ID NO. 82 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR15.

SEQ ID NO. 83 is the nucleotide sequence of the coding region of the mouse pheromone receptor VR16.

SEQ ID NO. 84 is the nucleotide sequence of the coding region of the rat pheromone receptor GoVN1.

SEQ ID NO. 85 is the nucleotide sequence of the coding region of the rat pheromone receptor GoVN2.

SEQ ID NO. 86 is the nucleotide sequence of the coding region of the rat pheromone receptor GoVN3.

SEQ ID NO. 87 is the nucleotide sequence of the coding region of the rat pheromone receptor GoVN4.

SEQ ID NO. 88 is the nucleotide sequence of the coding region of the rat pheromone receptor GoVN5.

SEQ ID NO. 89 is the nucleotide sequence of the coding region of the rat pheromone receptor GoVN6.

SEQ ID NO. 90 is the nucleotide sequence of the coding region of the rat pheromone receptor GoVN7.

SEQ ID NO. 91 is the nucleotide sequence of the coding region of the rat pheromone receptor GoVN13C.

SEQ ID NO. 92 is the nucleotide sequence of the coding region of the rat pheromone receptor GoVN13B.

Detailed Description of the Invention

The present invention in one aspect involves the cloning of cDNAs encoding several members of a multigene family of pheromone receptors. Complete cDNA sequences for selected murine and rat pheromone receptors are provided. Partial sequences of the human gene also are provided. The present invention also relates to the discovery that this family of pheromone receptors is expressed in a $G\alpha_o$ protein-expressing vomeronasal organ neurons ("$G\alpha_o^+$ VNO") or in another olfactory organ neuron in an animal (preferably, a mammal and more preferably, a human) which lacks a vomeronasal organ. Throughout this description, the pheromone receptors of the invention alternatively are referred to as "pheromone receptors", "$G\alpha_o^+$ VNO pheromone receptors" or, simply, "$G\alpha_o^+$ VNO receptors".

Analysis of the sequence homology between members of the receptor family by comparison to nucleic acid and protein databases established that the pheromone receptor family has several domains. These include, from amino terminus to carboxyl terminus:

(a) an amino-terminal extracellular domain containing from 30 to 600 amino acids; (b) a transmembrane region comprising: (i) seven non-contiguous transmembrane domains designated TM1, TM2, TM3, TM4, TM5, TM6 and TM7, (ii) three non-contiguous extracellular domains designated EC2, EC3 and EC4, and (iii) three non-contiguous intracellular domains designated IC1, IC2, and IC3, wherein the transmembrane domains, the extracellular domains and the intracellular domains are attached to one another from amino terminus to carboxyl terminus in the order TM1-IC1-TM2-EC2-TM3- IC2-TM4-EC3-TM5-IC3-TM6-EC4-TM7, and wherein the transmembrane region has at least about 35% homology and a length approximately equal to a transmembrane region of a polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 34, 36, 38, 40, 42, 44, 46, 48, and 50; and (c) a carboxyl-terminal intracellular domain containing from 5 to 200 amino acids. Each polypeptide member of the family is expressed in a $G\alpha_o$ protein-expressing vomeronasal organ neuron or are expressed in another olfactory organ neuron in an animal which does not possess a vomeronasal organ. One skilled in the art can readily identify olfactory organs in animals which do not possess a vomeronasal

organ. The homology can be calculated using various, publicly available software tools developed by NCBI (Bethesda, Maryland) that can be obtained through the internet (ftp://ncbi.nlm.nih.gov/pub/). Exemplary tools include the BLAST system. Pairwise and ClustalW alignments (BLOSUM30 matrix setting) as well as Kyte-Doolittle hydropathic analysis can be obtained using the MacVector sequence analysis software (Oxford Molecular Group).

The structure of the $G\alpha_o^+$ VNO pheromone receptors suggests that these receptors are members of the large G protein-coupled receptor superfamily (GPCR). Like other GPCRs, the $G\alpha_o^+$ VNO pheromone receptors exhibit seven hydrophobic stretches ("hydrophobic domains") and are similar in structure to other types of GPCRs, the calcium sensing receptor (CSR Ser. ID No. 59) and the metabotropic glutamate receptors (mGluRs). The CSR and mGluRs are unusual among the GPCRs in that they have extremely long N-terminal extracellular domain (e.g., 557-565 amino acids), a feature that is shared by the pheromone receptors of the invention. Despite this similarity, the receptors of the invention do not share substantial primary structure homology with the CSR and mGluRs. The receptors of the invention also are very different structurally from two other G-protein coupled receptors, the odorant receptors and $G\alpha i_2$+ vomeronasal receptors, which share none of the characteristic sequence motifs of the receptors of the invention and, moreover, which have very small (~12-28 amino acids) N-terminal extracellular domains.

The receptors of the invention differ somewhat in amino acid sequence, with regions of relatively high sequence homology. Refer to Examples 1 and 2 for a discussion and illustration of the amino acid sequence homology for the murine and rat $G\alpha_o^+$ VNO receptors, respectively. Other features of these members of the $G\alpha_o^+$ VNO receptor family also are discussed and illustrated in the Examples. For example, signal sequences have been identified for several of the $G\alpha_o^+$ VNO receptors disclosed in the Examples.

Homologs and alleles of the pheromone receptor nucleic acids of the invention can be identified by conventional techniques. Thus, an aspect of the invention is those nucleic acid sequences (SEQ ID NOs. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 54, and 55) which code for $G\alpha_o^+$ VNO pheromone receptors and which hybridize to a nucleic acid molecule consisting of the coding region of any one $G\alpha_o^+$ VNO pheromone receptor selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52, under high or low stringency conditions. The term "high or low stringency conditions" as used herein refers to parameters with which the art is familiar. Nucleic acid hybridization parameters may be found

in references which compile such methods, e.g. *Molecular Cloning: A Laboratory Manual*, J. Sambrook, et al., eds., Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, or *Current Protocols in Molecular Biology*, F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., New York. More specifically, high stringency conditions, as used herein, refers, for example, to hybridization at 65°C in hybridization buffer (3.5 x SSC, 0.02% Ficoll, 0.02% polyvinyl pyrolidone, 0.02% Bovine Serum Albumin, 2.5mM NaH_2PO_4(pH7), 0.5% SDS, 2mM EDTA). SSC is 0.15M sodium chloride/0.15M sodium citrate, pH7; SDS is sodium dodecyl sulphate; and EDTA is ethylenediaminetetracetic acid. Low stringency conditions would be the same, but with a lower temperature (e.g., 55°C). After hybridization, the membrane upon which the DNA is transferred is washed at 2 x SSC at room temperature and then at 0.2 x SSC/0.5% SDS at temperatures of up to 65°C. Additional conditions of varying stringency are provided in the Examples.

There are other conditions, reagents, and so forth which can used, which result in a similar degree of stringency. The skilled artisan will be familiar with such conditions, and thus they are not given here. It will be understood, however, that the skilled artisan will be able to manipulate the conditions in a manner to permit the clear identification of homologs and alleles of the $G\alpha_o^+$ VNO pheromone receptor nucleic acids of the invention. The skilled artisan also is familiar with the methodology for screening cells and libraries for expression of such molecules which then are routinely isolated, followed by isolation of the pertinent nucleic acid molecule and sequencing.

In general homologs and alleles typically will share at least 35% nucleotide identity and/or at least 50% amino acid identity to the cDNAs encoding a $G\alpha_o^+$ VNO pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52, in some instances will share at least 50% nucleotide identity and/or at least 65% amino acid identity and in still other instances will share at least 60% nucleotide identity and/or at least 75% amino acid identity. Watson-Crick complements of the foregoing nucleic acids also are embraced by the invention. As discussed above in the Summary of the invention, certain domains within the pheromone receptors may share even greater sequence homology to a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52.

In screening for $G\alpha_0^+$ VNO pheromone receptor polypeptides, a Southern blot may be performed using the foregoing conditions, together with a radioactive probe. After washing the membrane to which the DNA is finally transferred, the membrane can be placed against X-ray film to detect the radioactive signal.

The invention also includes degenerate nucleic acids which include alternative codons to those present in the native materials. For example, serine residues are encoded by the codons TCA, AGT, TCC, TCG, TCT and AGC. Each of the six codons is equivalent for the purposes of encoding a serine residue. Thus, it will be apparent to one of ordinary skill in the art that any of the serine-encoding nucleotide triplets may be employed to direct the protein synthesis apparatus, *in vitro* or *in vivo*, to incorporate a serine residue into an elongating $G\alpha_0^+$ VNO pheromone receptor polypeptide. Similarly, nucleotide sequence triplets which encode other amino acid residues include, but are not limited to,: CCA, CCC, CCG and CCT (proline codons); CGA, CGC, CGG, CGT, AGA and AGG (arginine codons); ACA, ACC, ACG and ACT (threonine codons); AAC and AAT (asparagine codons); and ATA, ATC and ATT (isoleucine codons). Other amino acid residues may be encoded similarly by multiple nucleotide sequences. Thus, the invention embraces degenerate nucleic acids that differ from the biologically isolated nucleic acids in codon sequence due to the degeneracy of the genetic code. In addition, areas of high similarity among pheromone receptors may differ in amino acid sequences such that they share many, but not all, amino acids. Their nucleotide sequences all differ accordingly.

The invention also provides isolated unique fragments of the cDNAs encoding a $G\alpha_0^+$ VNO polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52, or complements of these sequences. A unique fragment is one that is a 'signature' for the larger nucleic acid. It, for example, is long enough to assure that its precise sequence is not found in molecules outside of the $G\alpha_0^+$ VNO pheromone receptor nucleic acids defined above. Unique fragments can be used as probes in Southern blot assays to identify such nucleic acids, or can be used as primers in amplification assays such as those employing PCR. As known to those skilled in the art, large probes such as 200 nucleotides or more are preferred for certain uses such as Southern blots, while smaller fragments will be preferred for uses such as PCR. Unique fragments also can be used to produce fusion proteins for generating antibodies or determining binding of the polypeptide fragments, as demonstrated in the Examples, or for generating immunoassay

components. Likewise, unique fragments can be employed to produce nonfused fragments of the $G\alpha_o^+$ VNO pheromone receptor polypeptides, useful, for example, in the preparation of antibodies, in immunoassays, and as a competitive binding partner of the pheromones and/or other ligands which bind to the $G\alpha_o^+$ VNO pheromone receptor polypeptides, for example, in therapeutic applications. Unique fragments further can be used as antisense molecules to inhibit the expression of $G\alpha_o^+$ VNO pheromone receptor nucleic acids and polypeptides, particularly for the insecticide and other fertility control purposes as described in greater detail below.

As will be recognized by those skilled in the art, the size of the unique fragment will depend upon its conservancy in the genetic code. Thus, some regions of a cDNA selected from the group consisting of SEQ ID NO. 51, 53, 54, 55, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, and 92, that encodes a $G\alpha_o^+$ VNO polypeptide, and its complement will require longer segments to be unique while others will require only short segments, typically between 12 and 32 nucleotides (e.g. 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31 and 32 bases long). Virtually any segment of the region of the cDNAs encoding the full length $G\alpha_o^+$ VNO polypeptide or their complements, that is 18 or more nucleotides in length will be unique. Those skilled in the art are well versed in methods for selecting such sequences, typically on the basis of the ability of the unique fragment to selectively distinguish the sequence of interest from non-$G\alpha_o^+$ VNO pheromone receptor nucleic acids. A comparison of the sequence of the fragment to those on known data bases typically is all that is necessary, although *in vitro* confirmatory hybridization and sequencing analysis may be performed.

As mentioned above, the invention embraces antisense oligonucleotides that selectively bind to a nucleic acid molecule encoding a $G\alpha_o^+$ VNO pheromone receptor polypeptide, to decrease a pheromone receptor activity (e.g., a ligand binding activity, a signal transduction activity). This is desirable in virtually any condition wherein a reduction in pheromone binding or induction of a behavior that is triggered by pheromone binding is desirable, including to control fertility and behavior in vertebrates and invertebrates. The compositions of the invention are particularly useful in, for example, controlling fertility in livestock and controlling reproduction in rodents or insects by interrupting the normal behaviors of rodents or insects that result in reproduction. As used herein, the term "antisense oligonucleotide" or "antisense" describes an oligonucleotide that is an oligoribonucleotide, oligodeoxyribonucleotide, modified oligoribonucleotide, or modified oligodeoxyribonucleotide which hybridizes under physiological

conditions to DNA comprising a particular gene or to an mRNA transcript of that gene and, thereby, inhibits the transcription of that gene and/or the translation of that mRNA. The antisense molecules are designed so as to interfere with transcription or translation of a target gene upon hybridization with the target gene or transcript. Those skilled in the art will recognize that the exact length of the antisense oligonucleotide and its degree of complementarity with its target will depend upon the specific target selected, including the sequence of the target and the particular bases which comprise that sequence. It is preferred that the antisense oligonucleotide be constructed and arranged so as to bind selectively with the target under physiological conditions, i.e., to hybridize substantially more to the target sequence than to any other sequence in the target cell under physiological conditions. Based upon the cDNA sequences of Examples 1 and 2 (SEQ ID NOs. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 54, and 55), or upon allelic or homologous genomic and/or cDNA sequences, one of skill in the art can easily choose and synthesize any of a number of appropriate antisense molecules for use in accordance with the present invention. In order to be sufficiently selective and potent for inhibition, such antisense oligonucleotides should comprise at least 10 and, more preferably, at least 15 consecutive bases which are complementary to the target, although in certain cases modified oligonucleotides as short as 7 bases in length have been used successfully as antisense oligonucleotides (Wagner et al., *Nature Biotechnol.* 14:840-844, 1996). Most preferably, the antisense oligonucleotides comprise a complementary sequence of 20-30 bases. Although oligonucleotides may be chosen which are antisense to any region of the gene or mRNA transcripts, in preferred embodiments the antisense oligonucleotides correspond to N-terminal or 5' upstream sites such as translation initiation, transcription initiation or promoter sites. In addition, 3'-untranslated regions may be targeted. Targeting to mRNA splicing sites has also been used in the art but may be less preferred if alternative mRNA splicing occurs. In addition, the antisense is targeted, preferably, to sites in which mRNA secondary structure is not expected (see, e.g., Sainio et al., *Cell Mol. Neurobiol.* 14(5):439-457, 1994) and at which proteins are not expected to bind. Finally, although, Examples 1 and 2 disclose cDNA sequences (SEQ ID NOs. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 54, and 55), one of ordinary skill in the art may easily derive the genomic DNA corresponding to the cDNA of these cDNAs. Thus, the present invention also provides for antisense oligonucleotides which are complementary to the genomic DNA corresponding to a cDNA sequence selected from the group consisting of SEQ ID NOs. 1, 3, 5, 7, 9, 11, 13, 15, 17,

19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 54, and 55. Similarly, antisense to allelic or homologous cDNAs and genomic DNAs are enabled without undue experimentation.

In one set of embodiments, the antisense oligonucleotides of the invention may be composed of "natural" deoxyribonucleotides, ribonucleotides, or any combination thereof. That is, the 5' end of one native nucleotide and the 3' end of another native nucleotide may be covalently linked, as in natural systems, via a phosphodiester internucleoside linkage. These oligonucleotides may be prepared by art recognized methods which may be carried out manually or by an automated synthesizer. They also may be produced recombinantly by vectors.

In preferred embodiments, however, the antisense oligonucleotides of the invention also may include "modified" oligonucleotides. That is, the oligonucleotides may be modified in a number of ways which do not prevent them from hybridizing to their target but which enhance their stability or targeting or which otherwise enhance their therapeutic effectiveness.

The term "modified oligonucleotide" as used herein describes an oligonucleotide in which (1) at least two of its nucleotides are covalently linked via a synthetic internucleoside linkage (i.e., a linkage other than a phosphodiester linkage between the 5' end of one nucleotide and the 3' end of another nucleotide) and/or (2) a chemical group not normally associated with nucleic acids has been covalently attached to the oligonucleotide. Preferred synthetic internucleoside linkages are phosphorothioates, alkylphosphonates, phosphorodithioates, phosphate esters, alkylphosphonothioates, phosphoramidates, carbamates, carbonates, phosphate triesters, acetamidates, carboxymethyl esters and peptides.

The term "modified oligonucleotide" also encompasses oligonucleotides with a covalently modified base and/or sugar. For example, modified oligonucleotides include oligonucleotides having backbone sugars which are covalently attached to low molecular weight organic groups other than a hydroxyl group at the 3' position and other than a phosphate group at the 5' position. Thus modified oligonucleotides may include a 2'-O-alkylated ribose group. In addition, modified oligonucleotides may include sugars such as arabinose instead of ribose. The present invention, thus, contemplates pharmaceutical preparations containing modified antisense molecules that are complementary to and hybridizable with, under physiological conditions, nucleic acids encoding pheromone receptor polypeptides, together with pharmaceutically acceptable carriers.

Antisense oligonucleotides may be administered as part of a pharmaceutical composition. Such a pharmaceutical composition may include the antisense oligonucleotides in combination with any standard physiologically and/or pharmaceutically acceptable carriers which are known in the art. The compositions should be sterile and contain a therapeutically effective amount of the antisense oligonucleotides in a unit of weight or volume suitable for administration to a patient. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredients. The term "physiologically acceptable" refers to a non-toxic material that is compatible with a biological system such as a cell, cell culture, tissue, or organism. The characteristics of the carrier will depend on the route of administration. Physiologically and pharmaceutically acceptable carriers include diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials which are well known in the art.

As used herein, a "vector" may be any of a number of nucleic acids into which a desired sequence may be inserted by restriction and ligation for transport between different genetic environments or for expression in a host cell. Vectors are typically composed of DNA although RNA vectors are also available. Vectors include, but are not limited to, plasmids, phagemids and virus genomes. A cloning vector is one which is able to replicate in a host cell, and which is further characterized by one or more endonuclease restriction sites at which the vector may be cut in a determinable fashion and into which a desired DNA sequence may be ligated such that the new recombinant vector retains its ability to replicate in the host cell. In the case of plasmids, replication of the desired sequence may occur many times as the plasmid increases in copy number within the host bacterium or just a single time per host before the host reproduces by mitosis. In the case of phage, replication may occur actively during a lytic phase or passively during a lysogenic phase. An expression vector is one into which a desired DNA sequence may be inserted by restriction and ligation such that it is operably joined to regulatory sequences and may be expressed as an RNA transcript. Vectors may further contain one or more marker sequences suitable for use in the identification of cells which have or have not been transformed or transfected with the vector. Markers include, for example, genes encoding proteins which increase or decrease either resistance or sensitivity to antibiotics or other compounds, genes which encode enzymes whose activities are detectable by standard assays known in the art (e.g., ß-galactosidase or alkaline phosphatase), and genes which visibly affect the phenotype of transformed or transfected cells, hosts, colonies or plaques (e.g., green fluorescent protein).

Preferred vectors are those capable of autonomous replication and expression of the structural gene products present in the DNA segments to which they are operably joined.

As used herein, a coding sequence and regulatory sequences are said to be "operably" joined when they are covalently linked in such a way as to place the expression or transcription of the coding sequence under the influence or control of the regulatory sequences. If it is desired that the coding sequences be translated into a functional protein, two DNA sequences are said to be operably joined if induction of a promoter in the 5' regulatory sequences results in the transcription of the coding sequence and if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region to direct the transcription of the coding sequences, or (3) interfere with the ability of the corresponding RNA transcript to be translated into a protein. Thus, a promoter region would be operably joined to a coding sequence if the promoter region were capable of effecting transcription of that DNA sequence such that the resulting transcript might be translated into the desired protein or polypeptide.

The precise nature of the regulatory sequences needed for gene expression may vary between species or cell types, but shall in general include, as necessary, 5' non-transcribed and 5' non-translated sequences involved with the initiation of transcription and translation respectively, such as a TATA box, capping sequence, CAAT sequence, and the like. Especially, such 5' non-transcribed regulatory sequences will include a promoter region which includes a promoter sequence for transcriptional control of the operably joined gene. Regulatory sequences may also include enhancer sequences or upstream activator sequences as desired. The vectors of the invention may optionally include 5' leader or signal sequences. The choice and design of an appropriate vector is within the ability and discretion of one of ordinary skill in the art.

Expression vectors containing all the necessary elements for expression are commercially available and known to those skilled in the art. See, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989. Cells are genetically engineered by the introduction into the cells of heterologous DNA (RNA) encoding pheromone receptor polypeptide or fragment or variant thereof. That heterologous DNA (RNA) is placed under operable control of transcriptional elements to permit the expression of the heterologous DNA in the host cell.

Preferred systems for mRNA expression in mammalian cells are those such as pRc/CMV (available from Invitrogen, Carlsbad, CA) that contain a selectable marker such as a gene that

confers G418 resistance (which facilitates the selection of stably transfected cell lines) and the human cytomegalovirus (CMV) enhancer-promoter sequences. Additionally, suitable for expression in primate or canine cell lines is the pCEP4 vector (Invitrogen), which contains an Epstein Barr virus (EBV) origin of replication, facilitating the maintenance of plasmid as a multicopy extrachromosomal element. Another expression vector is the pEF-BOS plasmid containing the promoter of polypeptide Elongation Factor 1α, which stimulates efficiently transcription *in vitro*. The plasmid is described by Mishizuma and Nagata (*Nuc. Acids Res.* 18:5322, 1990), and its use in transfection experiments is disclosed by, for example, Demoulin (*Mol. Cell. Biol.* 16:4710-4716, 1996). Still another preferred expression vector is an adenovirus, described by Stratford-Perricaudet, which is defective for E1 and E3 proteins (*J. Clin. Invest.* 90:626-630, 1992). The use of the adenovirus as an Adeno.P1A recombinant is disclosed by Warnier et al., in intradermal injection in mice for immunization against P1A (*Int. J. Cancer*, 67:303-310, 1996).

The invention also embraces so-called expression kits, which allow the artisan to prepare a desired expression vector or vectors. Such expression kits include at least separate portions of each of the previously discussed coding sequences. Other components may be added, as desired, as long as the previously mentioned sequences, which are required, are included.

The invention also permits the construction of pheromone receptor gene "knock-outs" in cells and in animals, providing materials for studying certain aspects of pheromone receptor binding, signal transduction activity, or function.

The invention also provides isolated polypeptides, which include a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52 and unique fragments of these pheromone receptor polypeptides. Such polypeptides are useful, for example, alone or as fusion proteins to generate antibodies.

A unique fragment of a pheromone receptor polypeptide, in general, has the features and characteristics of unique fragments as discussed above in connection with nucleic acids. As will be recognized by those skilled in the art, the size of the unique fragment will depend upon factors such as whether the fragment constitutes a portion of a conserved protein domain. Thus, some regions of a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52

will require longer segments to be unique while others will require only short segments, typically between 5 and 12 amino acids (e.g. 5, 6, 7, 8, 9, 10, 11 and 12 amino acids long).

Unique fragments of a polypeptide preferably are those fragments which retain a distinct functional capability of the polypeptide. Functional capabilities which can be retained in a unique fragment of a polypeptide include interaction with antibodies, interaction with other polypeptides (G-proteins) or molecules (e.g., a ligand) or fragments thereof, selective binding of nucleic acids or proteins, and enzymatic activity. Those skilled in the art are well versed in methods for selecting unique amino acid sequences, typically on the basis of the ability of the unique fragment to selectively distinguish the sequence of interest from non-family members. A comparison of the sequence of the fragment to those on known data bases typically is all that is necessary.

The invention embraces variants of the pheromone receptor polypeptides described above. As used herein, a "variant" of a pheromone receptor polypeptide is a polypeptide which contains one or more modifications to the primary amino acid sequence of a pheromone receptor polypeptide. Modifications which create a pheromone receptor variant can be made to a pheromone receptor polypeptide 1) to reduce or eliminate an activity of a pheromone receptor polypeptide, such as a ligand binding activity or a signal transduction activity; 2) to enhance a property of a pheromone receptor polypeptide, such as protein stability in an expression system or the stability of protein-protein binding; or 3) to provide a novel activity or property to a pheromone receptor polypeptide, such as addition of an antigenic epitope or addition of a detectable moiety. Modifications to a pheromone receptor polypeptide are typically made to the nucleic acid which encodes the pheromone receptor polypeptide, and can include deletions, point mutations, truncations, amino acid substitutions and additions of amino acids or non-amino acid moieties. Alternatively, modifications can be made directly to the polypeptide, such as by cleavage, addition of a linker molecule, addition of a detectable moiety, such as biotin, addition of a fatty acid, and the like. Modifications also embrace fusion proteins comprising all or part of the pheromone receptor amino acid sequence.

In general, variants include pheromone receptor polypeptides which are modified specifically to alter a feature of the polypeptide unrelated to its physiological activity. For example, cysteine residues can be substituted or deleted to prevent unwanted disulfide linkages. Similarly, certain amino acids can be changed to enhance expression of a pheromone receptor polypeptide by eliminating proteolysis by proteases in an expression system.

Mutations of a nucleic acid which encode a pheromone receptor polypeptide preferably preserve the amino acid reading frame of the coding sequence, and preferably do not create regions in the nucleic acid which are likely to hybridize to form secondary structures, such a hairpins or loops, which can be deleterious to expression of the variant polypeptide.

Mutations can be made by selecting an amino acid substitution, or by random mutagenesis of a selected site in a nucleic acid which encodes the polypeptide. Variant polypeptides are then expressed and tested for one or more activities to determine which mutation provides a variant polypeptide with the desired properties. Further mutations can be made to variants (or to non-variant pheromone receptor polypeptides) which are silent as to the amino acid sequence of the polypeptide, but which provide preferred codons for translation in a particular host. The preferred codons for translation of a nucleic acid in, e.g., *E. coli*, are well known to those of ordinary skill in the art. Still other mutations can be made to the noncoding sequences of a pheromone receptor gene or cDNA clone to enhance expression of the polypeptide. The activity of variants of pheromone receptor polypeptides can be tested by cloning the gene encoding the variant pheromone receptor polypeptide into a bacterial or mammalian expression vector, introducing the vector into an appropriate host cell, expressing the variant pheromone receptor polypeptide, and testing for a functional capability of the pheromone receptor polypeptides as disclosed herein. For example, the variant pheromone receptor polypeptide can be tested for a ligand binding activity, wherein a ligand to which the receptor binds is contacted with the variant receptor and the amount of ligand binding to the variant receptor is determined using conventional procedures to measure the binding of one molecule to another. Preparation of other variant polypeptides may favor testing of other activities, as will be known to one of ordinary skill in the art.

The skilled artisan will also realize that conservative amino acid substitutions may be made in pheromone receptor polypeptides to provide functionally equivalent variants of the foregoing polypeptides, i.e, the variants retain the functional capabilities of the pheromone receptor polypeptides. As used herein, a "conservative amino acid substitution" refers to an amino acid substitution which does not alter the relative charge or size characteristics of the protein in which the amino acid substitution is made. Variants can be prepared according to methods for altering polypeptide sequence known to one of ordinary skill in the art such as are found in references which compile such methods, e.g. *Molecular Cloning: A Laboratory Manual*, J. Sambrook, et al., eds., Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring

Harbor, New York, 1989, or *Current Protocols in Molecular Biology*, F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., New York. To a certain extent, the various members of the pheromone receptor family that are illustrated in the Examples represent exemplary functionally equivalent variants of the pheromone receptor polypeptides. Other functionally equivalent variants include conservative amino acid substitutions of the amino acids of a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52. Conservative substitutions of amino acids include substitutions made amongst amino acids within the following groups: (a) M, I, L, V; (b) F, Y, W; (c) K, R, H; (d) A, G; (e) S, T; (f) Q, N; and (g) E, D.

Conservative amino-acid substitutions in the amino acid sequence of pheromone receptor polypeptides to produce functionally equivalent variants of pheromone receptor polypeptides typically are made by alteration of the nucleic acid encoding pheromone receptor polypeptides. Such substitutions can be made by a variety of methods known to one of ordinary skill in the art. For example, amino acid substitutions may be made by PCR-directed mutation, site-directed mutagenesis according to the method described in *Proc. Nat. Acad. Sci. U.S.A.* 82: 488-492, 1985, or by chemical synthesis of a gene encoding a pheromone receptor polypeptide. Where amino acid substitutions are made to a small unique fragment of a pheromone receptor polypeptide, such as a ligand binding site peptide, the substitutions can be made by directly synthesizing the peptide. The activity of functionally equivalent fragments of pheromone receptor polypeptides can be tested by cloning the gene encoding the altered pheromone receptor polypeptide into a bacterial or mammalian expression vector, introducing the vector into an appropriate host cell, expressing the altered pheromone receptor polypeptide, and testing for a functional capability of the pheromone receptor polypeptides as disclosed herein. Peptides which are chemically synthesized can be tested directly for function, e.g., for binding to a ligand to which the unaltered pheromone receptor is known to bind.

The invention as described herein has a number of uses, some of which are described elsewhere herein. First, the invention permits isolation of the pheromone receptor polypeptides of the Examples. A variety of methodologies well-known to the skilled practitioner can be utilized to obtain isolated pheromone receptor molecules. The polypeptide may be purified from cells which naturally produce the polypeptide by chromatographic means or immunological recognition. Alternatively, an expression vector may be introduced into cells to cause production of the polypeptide. In another method, mRNA transcripts may be microinjected or otherwise

introduced into cells to cause production of the encoded polypeptide. Translation of mRNA in cell-free extracts such as the reticulocyte lysate system also may be used to produce polypeptide. Those skilled in the art also can readily follow known methods for isolating pheromone receptor polypeptides. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography and immune-affinity chromatography.

The isolation of the pheromone receptor gene also makes it possible for the artisan to diagnose a disorder characterized by expression of pheromone receptor . These methods involve determining expression of the pheromone receptor gene, and/or pheromone receptor polypeptides derived therefrom. In the former situation, such determinations can be carried out via any standard nucleic acid determination assay, including the polymerase chain reaction as exemplified in the examples below, or assaying with labeled hybridization probes.

The invention also makes it possible to isolate the naturally occurring ligands (pheromones) and other ligands that have a ligand binding domain, namely, by the binding of such molecules to the pheromone receptor polypeptides (or fragments thereof containing a ligand binding site). Binding of the receptors to a ligand can be accomplished by introducing into a biological system in which the proteins bind (e.g., a cell) a molecule that includes a binding domain (putative ligand) in an amount sufficient to detect the binding.

The invention also provides agents such as binding polypeptides which bind to pheromone receptor polypeptides and/or to complexes of pheromone receptor polypeptides and their ligand binding partners. Such binding agents can be used, for example, in screening assays to detect the presence or absence of pheromone receptor polypeptides and complexes of pheromone receptor polypeptides and their ligand binding partners and in purification protocols to isolate pheromone receptor polypeptides and complexes of pheromone receptor polypeptides and their ligand binding partners. Such agents also can be used to inhibit the native activity of the pheromone receptor polypeptides or their ligand binding partners, for example, by binding to such polypeptides, or their binding partners or both.

The invention, therefore, embraces peptide binding agents which, for example, can be antibodies or fragments of antibodies having the ability to selectively bind to pheromone receptor polypeptides. Antibodies include polyclonal and monoclonal antibodies, prepared according to conventional methodology.

Significantly, as is well-known in the art, only a small portion of an antibody molecule, the paratope, is involved in the binding of the antibody to its epitope (see, in general, Clark, W.R. (1986) *The Experimental Foundations of Modern Immunology* Wiley & Sons, Inc., New York; Roitt, I. (1991) *Essential Immunology*, 7th Ed., Blackwell Scientific Publications, Oxford). The pFc' and Fc regions, for example, are effectors of the complement cascade but are not involved in antigen binding. An antibody from which the pFc' region has been enzymatically cleaved, or which has been produced without the pFc' region, designated an $F(ab')_2$ fragment, retains both of the antigen binding sites of an intact antibody. Similarly, an antibody from which the Fc region has been enzymatically cleaved, or which has been produced without the Fc region, designated an Fab fragment, retains one of the antigen binding sites of an intact antibody molecule. Proceeding further, Fab fragments consist of a covalently bound antibody light chain and a portion of the antibody heavy chain denoted Fd. The Fd fragments are the major determinant of antibody specificity (a single Fd fragment may be associated with up to ten different light chains without altering antibody specificity) and Fd fragments retain epitope-binding ability in isolation.

Within the antigen-binding portion of an antibody, as is well-known in the art, there are complementarity determining regions (CDRs), which directly interact with the epitope of the antigen, and framework regions (FRs), which maintain the tertiary structure of the paratope (see, in general, Clark, 1986; Roitt, 1991). In both the heavy chain Fd fragment and the light chain of IgG immunoglobulins, there are four framework regions (FR1 through FR4) separated respectively by three complementarity determining regions (CDR1 through CDR3). The CDRs, and in particular the CDR3 regions, and more particularly the heavy chain CDR3, are largely responsible for antibody specificity.

It is now well-established in the art that the non-CDR regions of a mammalian antibody may be replaced with similar regions of nonspecific or heterospecific antibodies while retaining the epitopic specificity of the original antibody. This is most clearly manifested in the development and use of "humanized" antibodies in which non-human CDRs are covalently joined to human FR and/or Fc/pFc' regions to produce a functional antibody. Thus, for example, PCT International Publication Number WO 92/04381 teaches the production and use of humanized murine RSV antibodies in which at least a portion of the murine FR regions have been replaced by FR regions of human origin. Such antibodies, including fragments of intact antibodies with antigen-binding ability, are often referred to as "chimeric" antibodies.

Thus, as will be apparent to one of ordinary skill in the art, the present invention also provides for $F(ab')_2$, Fab, Fv and Fd fragments; chimeric antibodies in which the Fc and/or FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; chimeric $F(ab')_2$ fragment antibodies in which the FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; chimeric Fab fragment antibodies in which the FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; and chimeric Fd fragment antibodies in which the FR and/or CDR1 and/or CDR2 regions have been replaced by homologous human or non-human sequences. The present invention also includes so-called single chain antibodies.

Thus, the invention involves polypeptides of numerous size and type that bind specifically to pheromone receptor polypeptides, and/or complexes of both pheromone receptor polypeptides and their ligand binding partners. These polypeptides may be derived also from sources other than antibody technology. For example, such polypeptide binding agents can be provided by degenerate peptide libraries which can be readily prepared in solution, in immobilized form or as phage display libraries. Combinatorial libraries also can be synthesized of peptides containing one or more amino acids. Libraries further can be synthesized of peptoids and non-peptide synthetic moieties.

Phage display can be particularly effective in identifying binding peptides useful according to the invention. Briefly, one prepares a phage library (using e.g. m13, fd, or lambda phage), displaying inserts from 4 to about 80 amino acid residues using conventional procedures. The inserts may represent, for example, a completely degenerate or biased array. One then can select phage-bearing inserts which bind to the pheromone receptor polypeptide. This process can be repeated through several cycles of reselection of phage that bind to the pheromone receptor polypeptide. Repeated rounds lead to enrichment of phage bearing particular sequences. DNA sequence analysis can be conducted to identify the sequences of the expressed polypeptides. The minimal linear portion of the sequence that binds to the pheromone receptor polypeptide can be determined. One can repeat the procedure using a biased library containing inserts containing part or all of the minimal linear portion plus one or more additional degenerate residues upstream or downstream thereof. Yeast two-hybrid screening methods also may be used to identify polypeptides that bind to the pheromone receptor polypeptides. Thus, the pheromone receptor polypeptides of the invention, or a fragment thereof, can be used to screen peptide

libraries, including phage display libraries, to identify and select peptide binding partners of the pheromone receptor polypeptides of the invention. Such molecules can be used, as described, for screening assays, for purification protocols, for interfering directly with the functioning of pheromone receptor and for other purposes that will be apparent to those of ordinary skill in the art.

A pheromone receptor polypeptide, or a fragment which contains the ligand binding site, also can be used to isolate naturally-occurring ligands and other binding partners of the receptors of the invention. For example, an isolated pheromone receptor can be used to isolate ligands that bind to the receptor binding site by immobilizing a receptor (or fragment containing the ligand binding site) on a chromatographic media, such as polystyrene beads, or a filter, and using the immobilized polypeptide to isolate molecules that bind to this affinity matrix in accordance with standard procedures for affinity chromatography.

It will also be recognized that the invention embraces the use of the pheromone receptor cDNA sequences in expression vectors, as well as to transfect host cells and cell lines, be these prokaryotic (e.g., *E. coli*), or eukaryotic (e.g., CHO cells, COS cells, yeast expression systems and recombinant baculovirus expression in insect cells). Especially useful are oocytes, mammalian cells such as mouse, hamster, pig, goat, primate, etc. They may be of a wide variety of tissue types, and include primary cells and cell lines. The expression vectors require that the pertinent sequence, i.e., those nucleic acids described *supra*, be operably linked to a promoter.

When administered, the therapeutic compositions of the present invention are administered in pharmaceutically acceptable preparations. Such preparations may routinely contain pharmaceutically acceptable concentrations of salt, buffering agents, preservatives, compatible carriers, supplementary immune potentiating agents such as adjuvants and cytokines and optionally other therapeutic agents.

The therapeutics of the invention can be administered by any conventional route, including injection or by gradual infusion over time. The administration may, for example, be oral, intravenous, intraperitoneal, intramuscular, intracavity, subcutaneous, or transdermal. When antibodies are used therapeutically, a preferred route of administration is by pulmonary aerosol. Techniques for preparing aerosol delivery systems containing antibodies are well known to those of skill in the art. Generally, such systems should utilize components which will not significantly impair the biological properties of the antibodies, such as the paratope binding

capacity (see, for example, Sciarra and Cutie, "Aerosols," in Remington's Pharmaceutical Sciences, 18th edition, 1990, pp 1694-1712; incorporated by reference). Those of skill in the art can readily determine the various parameters and conditions for producing antibody aerosols without resort to undue experimentation. When using antisense preparations of the invention, slow intravenous administration is preferred.

Preparations for parenteral administration include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's or fixed oils. Intravenous vehicles include fluid and nutrient replenishers, electrolyte replenishers (such as those based on Ringer's dextrose), and the like. Preservatives and other additives may also be present such as, for example, antimicrobials, anti-oxidants, chelating agents, and inert gases and the like.

The preparations of the invention are administered in effective amounts. An effective amount is that amount of a pharmaceutical preparation that alone, or together with further doses, produces the desired response in the condition being treated, e.g., modifying fertility or pheromone-mediated behaviors that are related to reproduction or aggression. For example, this can involve the use of the compounds of the invention as pesticides to slow or halt insect or rodent behaviors that result in reproduction. Alternatively, this can involve the use of the compounds of the invention as agents for controlling fertility in animals (e.g., livestock, domestic animals), by providing compounds which inhibit or stimulate the behaviors in such animals that result in reproduction or agression. This can be monitored by routine methods, e.g., observing the behavior in the animal (vertebrate or invertebrate) recipient.

The invention also contemplates gene therapy, e.g., to prepare an animal model for studying the conditions and behaviors (e.g., fertility, aggression) that are pheromone receptor-mediated. The procedure for performing *ex vivo* gene therapy is outlined in U.S. Patent 5,399,346 and in exhibits submitted in the file history of that patent, all of which are publicly available documents. In general, it involves introduction *in vitro* of a functional copy of a gene into a cell(s) of a subject which contains a defective copy of the gene, and returning the genetically engineered cell(s) to the subject. The functional copy of the gene is under operable control of regulatory elements which permit expression of the gene in the genetically engineered

cell(s). Numerous transfection and transduction techniques as well as appropriate expression vectors are well known to those of ordinary skill in the art, some of which are described in PCT application WO95/00654. *In vivo* gene therapy using vectors such as adenovirus, retroviruses, herpes virus, and targeted liposomes also is contemplated according to the invention.

The invention further provides efficient methods of identifying pharmacological agents or lead compounds for agents active at the level of a pheromone receptor or pheromone receptor fragment modulatable cellular function. In particular, such functions include ligand binding activity. Generally, the screening methods involve assaying for activation of pheromone receptors or assaying for compounds which interfere with a pheromone receptor activity such as pheromone receptor binding to its cognate ligand. Such methods are adaptable to automated, high throughput screening of compounds. The target therapeutic indications for pharmacological agents detected by the screening methods that block pheromone receptor activity are limited only in that the target cellular function be subject to modulation by alteration of the formation of a complex comprising a pheromone receptor polypeptide or fragment thereof and one or more natural pheromone receptor ligands. Target indications include cellular processes modulated by pheromone receptor signal transduction following receptor-ligand binding.

A wide variety of assays for pharmacological agents are provided, including, labeled *in vitro* protein-protein binding assays, electrophoretic mobility shift assays, immunoassays, cell-based assays such as two- or three-hybrid screens, expression assays, activation of G-proteins, etc. For example, three-hybrid screens are used to rapidly examine the effect of transfected nucleic acids on the intracellular binding of pheromone receptor or pheromone receptor fragments to specific extracellular targets (e.g., ligands in biological samples, such as urine, vaginal fluid, or in combinatorial libraries) .

Pheromone receptor fragments used in the methods, when not produced by a transfected nucleic acid are added to an assay mixture as an isolated polypeptide. The assay can be used to screen putative ligands for their ability to bind to the receptor. Pheromone receptor polypeptides preferably are produced recombinantly, although such polypeptides may be isolated from biological extracts. Recombinantly produced pheromone receptor polypeptides include chimeric proteins comprising a fusion of a pheromone receptor protein with another polypeptide. For example, a polypeptide fused to a pheromone receptor polypeptide or fragment may also provide means of readily detecting the fusion protein, e.g., by immunological recognition or by fluorescent labeling.

In addition to the pheromone receptor, a screening assay mixture includes a binding partner for the receptor, e.g., a naturally occurring ligand that is capable of binding to the pheromone receptor or, alternatively, is comprised of an analog which mimics the pheromone receptor binding properties of the naturally occurring ligand for purposes of the assay. The screening assay mixture also comprises a candidate pharmacological agent (e.g., a putative receptor agonist or antagonist). Typically, a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a different response to the various concentrations. Typically, one of these concentrations serves as a negative control, i.e., at zero concentration of agent or at a concentration of agent below the limits of assay detection. Candidate agents encompass numerous chemical classes, although typically they are organic compounds. Preferably, the candidate pharmacological agents are small organic compounds, i.e., those having a molecular weight of more than 50 yet less than about 2500, preferably less than about 1000 and, more preferably, less than about 500. Candidate agents comprise functional chemical groups necessary for structural interactions with polypeptides and/or nucleic acids, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups and more preferably at least three of the functional chemical groups. The candidate agents can comprise cyclic carbon or heterocyclic structure and/or aromatic or polyaromatic structures substituted with one or more of the above-identified functional groups. Candidate agents also can be biomolecules such as peptides, saccharides, fatty acids, sterols, isoprenoids, purines, pyrimidines, derivatives or structural analogs of the above, or combinations thereof and the like. Where the agent is a nucleic acid, the agent typically is a DNA or RNA molecule, although modified nucleic acids as defined herein are also contemplated.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides, synthetic organic combinatorial libraries, phage display libraries of random peptides, and the like. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural and synthetically produced libraries and compounds can be readily be modified through conventional chemical, physical, and biochemical means. Further, known pharmacological agents may be subjected to directed or random chemical modifications such as

acylation, alkylation, esterification, amidification, etc. to produce structural analogs of the agents.

A variety of other reagents also can be included in the mixture. These include reagents such as salts, buffers, neutral proteins (e.g., albumin), detergents, etc. which may be used to facilitate optimal protein-protein and/or protein-nucleic acid binding. Such a reagent may also reduce non-specific or background interactions of the reaction components. Other reagents that improve the efficiency of the assay such as protease, inhibitors, nuclease inhibitors, antimicrobial agents, and the like may also be used.

The mixture of the foregoing assay materials is incubated under conditions whereby, but for the presence of the candidate pharmacological agent, the pheromone receptor polypeptide specifically binds the cellular binding target, a portion thereof or analog thereof. The order of addition of components, incubation temperature, time of incubation, and other parameters of the assay may be readily determined. Such experimentation merely involves optimization of the assay parameters, not the fundamental composition of the assay. Incubation temperatures typically are between 4°C and 40°C. Incubation times preferably are minimized to facilitate rapid, high throughput screening, and typically are between 0.1 and 10 hours.

After incubation, the presence or absence of specific binding between the pheromone receptor polypeptide and one or more binding targets is detected by any convenient method available to the user. For cell free binding type assays, a separation step is often used to separate bound from unbound components. The separation step may be accomplished in a variety of ways. Conveniently, at least one of the components is immobilized on a solid substrate, from which the unbound components may be easily separated. The solid substrate can be made of a wide variety of materials and in a wide variety of shapes, e.g., microtiter plate, microbead, dipstick, resin particle, etc. The substrate preferably is chosen to maximum signal to noise ratios, primarily to minimize background binding, as well as for ease of separation and cost.

Separation may be effected for example, by removing a bead or dipstick from a reservoir, emptying or diluting a reservoir such as a microtiter plate well, rinsing a bead, particle, chromatographic column or filter with a wash solution or solvent. The separation step preferably includes multiple rinses or washes. For example, when the solid substrate is a microtiter plate, the wells may be washed several times with a washing solution, which typically includes those components of the incubation mixture that do not participate in specific bindings such as salts,

buffer, detergent, non-specific protein, etc. Where the solid substrate is a magnetic bead, the beads may be washed one or more times with a washing solution and isolated using a magnet.

Detection may be effected in any convenient way for cell-based assays such as two- or three-hybrid screens. The transcript resulting from a reporter gene transcription assay of Pheromone receptor polypeptide binding to a target molecule typically encodes a directly or indirectly detectable product, e.g., β-galactosidase activity, luciferase activity, and the like. A wide variety of cell based assays for G-protein coupled receptors could also be employed for detection of molecules that stimulate (agonsists) pheromone receptors or block (agonists) that stimulation by natural ligands or agonists. Pheromone receptor polypeptides or chimeric receptors composed only in-part of a pheromone receptor could be employed in these assays. The chimeric receptors might, for example, contain part of another G-protein coupled receptor such that binding of a ligand to the pheromone receptor binding domain results in coupling to a particular G-protein where activation could be easily assayed. For cell free binding assays, one of the components usually comprises, or is coupled to, a detectable label. A wide variety of labels can be used, such as those that provide direct detection (e.g., radioactivity, luminescence, optical or electron density, etc). or indirect detection (e.g., epitope tag such as the FLAG epitope, enzyme tag such as horseradish peroxidase, etc.). The label may be bound to a pheromone receptor binding partner (ligand), or incorporated into the structure of the binding partner.

A variety of methods may be used to detect the label, depending on the nature of the label and other assay components. For example, the label may be detected while bound to the solid substrate or subsequent to separation from the solid substrate. Labels may be directly detected through optical or electron density, radioactive emissions, nonradioactive energy transfers, etc. or indirectly detected with antibody conjugates, strepavidin-biotin conjugates, etc. Methods for detecting the labels are well known in the art.

The invention provides pheromone receptor -specific binding agents, methods of identifying and making such agents, and their use in diagnosis, therapy and pharmaceutical development, including the development of pesticides and other agents for controlling fertility and reproduction (or related behaviors) in animals. For example, pheromone receptor-specific pharmacological agents are useful in a variety of diagnostic and therapeutic applications, especially where disease or disease prognosis is associated with improper utilization of a pathway involving pheromone receptor. Novel pheromone receptor-specific binding agents include pheromone receptor-specific antibodies and other natural intracellular binding agents

identified with assays such as two hybrid screens, and non-natural intracellular binding agents identified in screens of chemical libraries and the like.

In general, the specificity of pheromone receptor binding to a binding agent is shown by binding equilibrium constants. Targets which are capable of selectively binding a pheromone receptor polypeptide preferably have binding equilibrium constants of at least about 10^7 M^{-1}, more preferably at least about 10^8 M^{-1}, and most preferably at least about 10^9 M^{-1}. The wide variety of cell based and cell free assays may be used to demonstrate pheromone receptor -specific binding. Cell based assays include one, two and three hybrid screens, assays in which pheromone receptor -mediated transcription is inhibited or increased activation of G-proteins, etc. Cell free assays include pheromone receptor -protein binding assays, immunoassays, etc. Other assays useful for screening agents which bind pheromone receptor polypeptides include fluorescence resonance energy transfer (FRET), and electrophoretic mobility shift analysis (EMSA).

Various techniques may be employed for introducing nucleic acids of the invention into cells, depending on whether the nucleic acids are introduced *in vitro* or *in vivo* in a host. Such techniques include transfection of nucleic acid-$CaPO_4$ precipitates, transfection of nucleic acids associated with DEAE, transfection with a retrovirus including the nucleic acid of interest, liposome mediated transfection, and the like. For certain uses, it is preferred to target the nucleic acid to particular cells. In such instances, a vehicle used for delivering a nucleic acid of the invention into a cell (e.g., a retrovirus, or other virus; a liposome) can have a targeting molecule attached thereto. For example, a molecule such as an antibody specific for a surface membrane protein on the target cell or a ligand for a receptor on the target cell can be bound to or incorporated within the nucleic acid delivery vehicle. For example, where liposomes are employed to deliver the nucleic acids of the invention, proteins which bind to a surface membrane protein associated with endocytosis may be incorporated into the liposome formulation for targeting and/or to facilitate uptake. Such proteins include capsid proteins or fragments thereof tropic for a particular cell type, antibodies for proteins which undergo internalization in cycling, proteins that target intracellular localization and enhance intracellular half life, and the like. Polymeric delivery systems also have been used successfully to deliver nucleic acids into cells, as is known by those skilled in the art. Such systems even permit oral delivery of nucleic acids.

Examples

Example 1

Experimental Procedures

Preparation and analysis of single cell cDNAs

Male mouse (C57BL/6J) VNOs were minced, incubated in Trypsin-EDTA (Gibco-BRL/LTI, Rockville, Maryland), and triturated to obtain dissociated cells. The cells were centrifuged (1000 RPM, 5 min) and resuspended in phosphate buffered saline + 0.1% bovine serum albumin. Individual cells that appeared to be neurons were transferred to separate tubes with a microcapillary pipet.

cDNAs were prepared from each cell and amplified according to Brady and Iscove (*Methods in Enzymology*, 1993, 225:611-621) with minor modifications. Briefly, cDNAs were prepared from the 3' ends of mRNAs by reverse transcription with an oligo (dT) primer, and a poly dA stretch was added to each cDNA with terminal transferase. The cDNAs were then amplified by PCR with one of two primers, AL1 (ATTGGATCCAGGCCGCTCTGGACAA AATATGAA TTC(T) (SEQ. ID. No. 56) (Dulac and Axel, *Cell*, 1995, 83:195-206 or AL3 (GGCACATGG ACGAAATCTTGGTACTCTTCAGAATTC(T), (SEQ. ID. No. 57) and Taq polymerase [Amplitaq LD ("ALD") or Amplitaq Stoffel Fragment ("ASF") (Perkin Elmer, Norwalk, CT)].

Aliquots of each cDNA sample were electrophoresed on agarose gels and blotted onto nylon membranes (Hybond N^+, Amersham, Piscataway, NJ) (Ausubel, F., et al., *Current Protocols in Molecular Biology*, 1988, John Wiley & Sons NY, NY; Sambrook, J., et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989). The blots were hybridized at 55° or 70°C in Hyb Buffer (0.5M sodium phosphate buffer (pH7.3), 4% SDS, 1% bovine serum albumin (BSA)) with ^{32}P-labeled probes prepared by random priming (Prime-It II, Stratagene, La Jolla, CA).

Construction and screening of single cell cDNA libraries

An aliquot of cDNA sample VN14 was digested with Eco RI and gel-isolated fragments of 0.1-1.5 kb were cloned into λZapII Ausubel, F., et al., *Current Protocols in Molecular Biology*, 1988, John Wiley & Sons NY, NY; Sambrook, J., et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989). Two

thousand library clones were plated at low density. Replica filter lifts were hybridized at 75°C (in Hyb Buffer containing 2μg/ml poly (dT)24 and 1μg/ml of random dA-dT 20-mers) to ^{32}P-labeled probes (~2.5 x 10^8 CPM/μg; 5 x 10^6 CPM/ml) prepared by PCR of different single cell cDNA samples. Clones that hybridized to only a VN14 probe were isolated, and a probe prepared from the insert of each was hybridized to blots of selected single cell cDNAs. Clones that hybridized to only VN14 cDNAs were sequenced.

Isolation and analysis of VR cDNA clones

sc153, one VN14$^+$VN2$^-$ clone from the VN14 library, was used as probe to screen a mouse VNO cDNA library ('λVNO') (Berghard, A., et al., *J Neurosci*, 1996, 16:909-918) and a mouse genomic DNA library (Stratagene, La Jolla, CA) (70°C, Hyb buffer). Hybridizing clones were found only in the genomic library. A fragment containing 2kb upstream of sc153 was isolated from one genomic clone (153G1) and used to screen lVNO (55°C, Hyb Buffer). The region (D10-TM7) of one clone (D10) that showed homology to TM7 of the CSR (SEQ ID NO. 59) was then used to screen lVNO (55°C, Hyb Buffer), yielding a variety of VR cDNA clones. Additional clones were obtained from lVNO using probes prepared from clones previously isolated, or from PCR products obtained by amplification of mouse genomic DNA or VNO cDNA with degenerate primers (Buck, L., et al., *Cell*, 1991, 65:175-187) matching conserved motifs in the VRs. Some PCR products were also cloned into pCR2.1 (Invitrogen, Carlsbad, CA) and sequenced.

Analysis of VR mRNAs by RT-PCR

Random-primed cDNA prepared from male or female C57BL/6J mouse VNO RNAs (or VR cDNA clones) were used in PCR reactions with degenerate primers (Buck and Axel, *Cell* 1991, 65:175-187) matching conserved VR motifs to amplify VR sequences corresponding to amino acids 33-772 in VR1 (SEQ ID NO. 2). Nested PCR was performed with a 1/1000 dilution of the first PCR reaction and primer pairs matching regions of putative exons 1 and 6 in specific VR cDNA clones. Blots prepared from size-fractionated, nested PCR products were hybridized (70°C, Hyb buffer containing 100μg/ml herring sperm DNA (Sigma, St Louis, MO)) to probes prepared from the PCR products of the cDNA clones.

Northern and S uthern blots and gen mic library screens

Northern Blots: One μg of PolyA$^+$ RNA prepared from mouse VNO and OE, or purchased from Clontech (other tissue RNAs), was size fractionated on formaldehyde gels, and blotted (see above) (Berghard and Buck, *J Neurosci*, 1996, 16:909-918). The blot was hybridized (70°C, Hyb Buffer) with a ^{32}P-labeled probe prepared from the regions of cDNAs VR1, VR2, VR4, and VR15 corresponding to that encoding amino acids 33-772 in VR1 (SEQ ID NO. 1).

Southern Blots: 5 μg of genomic DNA prepared from C57BL6/J mouse liver was digested with Eco RI or Hind III, size fractionated, and blotted (Ressler et al, *Cell*, 1993, 73:597-609). The blots were hybridized (70°C, Hyb buffer containing sperm DNA (see above)) to probes prepared from 3' untranslated segments of different VR cDNA clones [VR2 (nt.2607-2961 of SEQ ID NO. 3), VR3 (nt. 2505-2907 of SEQ ID NO. 5), and VR15 (nt. 3239-3689 of SEQ ID NO. 29)]. A VR4 probe was also used, which gave the same results as highly related VR15 probe.

Genomic library screens to determine VR gene number: A mouse genomic library was screened separately at 70°C or 55°C (see above) with different ^{32}P-labeled probes. Probe 1: a mix of segments of cDNAs VR1 (SEQ ID NO. 1), VR2 (SEQ ID NO. 3), VR4 (SEQ ID NO. 7), and VR15 (SEQ ID NO. 29) encoding the region corresponding to amino acids 619-772 of VR1 (SEQ ID NO. 2). Probes 2-6: Segments of VR genes obtained from mouse genomic DNA by PCR with degenerate primers matching conserved VR sequence motifs. The PCR segments corresponded to the following amino stretches in VR1 (SEQ ID NO. 2): amino acids 191-397, 565-825, 637-825, 637-804, and 619-784. For example, degenerate oligonucleotide primer pairs used included:

for amino acids 191-397:

5' primer= (GCT)TI(CT)A(CT) CA(AG)(AG)TIGCI(AC)CIAA(AG)GA(CT)AC (SEQ ID NO. 60),

3' primer= G(CT)(AG)T(GT)IGCI(AG)(CT)I(AG)C(AG)T(AG)IACI(AG)C(AG)TT (SEQ ID NO. 61);

for amino acids 565-825:

5' primer= (AC)(AG)ITG (CT)CCI(GT)AIIA(CT)(AC)A(AG)TA(CT)GCIAA (SEQ ID NO. 62),

3' primer= GIC(GT)IA(CT)IA(AG)IATIA (CT)(AG)TAI(AC)(AT)(CT)TTIGGIAC (SEQ ID NO. 63);

for amino acids 637-825:

5' primer= ATI(AT)(GC)I (CT) TI(AG)TITT(CT)TG(CT)TT(CT)(CT)TITG (SEQ ID NO. 64),

3' primer= GIC(GT)IA(CT)IA(AG)IATIA (CT)(AG)TAI(AC)(AT)(CT)TTIGGIAC (SEQ ID NO. 63);

for amino acids 637-804:

5' primer= ATI(AT)(GC)I(CT)TI(AG)TITT(CT)TG(CT)TT(CT)(CT)TITG (SEQ ID NO. 64),

3' primer= (AG)IATI(GC)(AT)(AG)AAIA(CT)(CT)TCIACI (AG)CIACCAT (SEQ ID NO. 65); and

for amino acids 619-784:

5' primer= GA(CT)ACICCIATIGTIAA(AG)GCIAA(CT)AA (SEQ ID NO. 66),

3' primer= AAIGTIA(CT)CCAIACI(GC)(AT)(AG)CA(AG)AAIAC (SEQ ID NO. 67), wherein all primers are in a 5'→3' direction, I:Inosine.

In situ hybridization

In situ hybridization was performed according to Schaeren-Wiemers and Gerfin-Moser (*Histochemistry*, 1993, 100:431-440) with sequential 16 micron sections of male or female VNOs. Digoxigenin- labeled cRNA probes were prepared from the same 3' untranslated regions of VR cDNAs as used for the genomic Southern blots. Sections were counter-stained with Hoechst 33258, which labels nuclei. The numbers of $G_{\alpha o}$- or $G_{\alpha i2}$-labeled cells (or cells labeled with VR probes) was determined by counting the number of nuclei in labeled regions. The total number of cells was considered to be the sum of $G_{\alpha o}$+ and $G_{\alpha i2}$+ cells in adjacent sections.

Chromosome mapping of VR genes

Southern blots of genomic DNA from C57BL/6J and Mus spretus (Jackson Labs) digested with different restriction enzymes were prepared and probed with specific VR cDNA probes as described above. Southern blots of Eco RI, size fractionated genomic DNAs from 94 different backcross mice (M. spretus x (M. spretus x C57BL/6J)), were purchased from Jackson Labs. These blots were hybridized to probes prepared from 3' untranslated segments of the VR2 or VR4 (see above) cDNA at 70°C and washed (see above). Polymorphic bands were typed as either M. spretus or M. spretus/C57BL/6J. The data was sent to the Jackson Laboratory Backcross DNA Mapping Panel Resource for determination of the chromosomal locations of the

polymorphic fragments. Additional information was obtained via internet from Jackson Laboratory Mouse Genome Informatics.

Cloning of a gene differentially expressed in G_{ao}+ VNs

Different members of the OR and VNR families are expressed in different neurons in the OE and $G_{\alpha i2}$+ zone of the VNO, respectively. It therefore appeared likely that the same would be true of sensory receptors expressed by G_{ao}+ VNs. The differential screening of cDNA libraries with cDNA probes prepared from a few neurons can be used to identify genes expressed in one neuron, but not another (Buck, L., et al, *Annu. Rev. Neurosci.*, 1996, 19:517-544). Using PCR, this can be accomplished with single cells (Brady, G., et al., *Methods in Enzymology*, 1993, 225:611-621; Dulac, C., et al., *Cell*, 1995, 83:195-206).

To search for genes encoding receptors expressed by G_{ao}+ VNs, we looked for genes expressed in one G_{ao}+ VN, but not another, using the PCR-based differential screening approach. In initial experiments, we isolated a series of mouse VNs, prepared cDNAs from the 3' ends of mRNAs present in each, and amplified the single-cell cDNA fragments by PCR. Many of the amplified, single-cell cDNA samples hybridized to an OMP probe, confirming their derivation from VNs (Berghard et al, *Proc. Natl. Acad. Sci. USA,* 1996, 93:2365-2369). With one exception, G_{ao} and $G_{\alpha i2}$ probes hybridized to different OMP+ samples, allowing us to identify samples that were derived from G_{ao}+ VNs.

We next prepared a library from one of the G_{ao}+ single-cell cDNA samples (VN14), and isolated clones that hybridized to a probe prepared from VN14, but not to a probe prepared from another G_{ao}+ sample (VN2). We identified 3 VN14+VN2- clones, which differed in size, but were otherwise identical in sequence. None contained an open reading frame, which was not surprising since, in the method used, the amplified cDNAs are only ~400-800 bp long, and are derived from the 3' ends of mRNAs (Brady and Iscove, *Methods in Enzymology,* 1993, 225:611-621).

We next hybridized one of the VN14+VN2- clones (sc153) to the original panel of single-cell cDNAs. sc153 hybridized to VN14, but not to any of the other cDNA samples. Consistent with this result, sc153 hybridized to only a small percentage (~0.3%) of VNs in VNO tissue sections.

Using sc153 as probe, we were able to isolate a sc153+ clone from a mouse genomic library which contained ~2 kb of DNA 5' to the sc153 sequence. Using this 2kb fragment as

probe, we isolated a matching clone (D10) from the VNO cDNA library. Sequence analysis showed that sc153 and D10 were derived from the same gene, but that the D10 cDNA was truncated at the 3' end and did not contain the final 685 bp of sequence present in sc153. Like sc153, D10 hybridized to only a small percentage of VNs in VNO tissue sections.

The 5' end of the D10 cDNA contained a short open reading frame, which encoded a protein fragment with homology to transmembrane domain 7 (TM7) of the calcium sensing receptor (CSR), a G protein-coupled receptor (GPCR) (Brown et al, *Nature*, 1993, 366:575-580). When the TM7-related region of D10 (D10-TM7) was hybridized at reduced stringency (55°C) to the original panel of single-cell cDNAs, it labeled many of the $G_{\alpha o}$+ samples, but none of $G_{\alpha i2}$+ ones (except the one that was also $G_{\alpha o}$+, and was probably derived from two cells). Since D10 labeled only a small percentage of VNs in tissue sections under high stringency conditions, this suggested that many $G_{\alpha o}$+ neurons express a gene related to D10, but not identical to it.

A novel multigene family encoding VNO receptors

Hybridization of D10-TM7 to the VNO cDNA library at reduced stringency yielded a number of related cDNA clones (e.g. VR1-VR3, SEQ ID NOs. 1-6). Additional related cDNAs were obtained by RT-PCR with degenerate primers (e.g. VR6-VR7, SEQ ID NOs. 11-14), or by screening the VNO cDNA library with a PCR product obtained from genomic DNA (e.g., VR4, VR5, SEQ ID NOs. 7-10).

These cDNAs encode a novel family of proteins, which are members of the G protein-coupled receptor (GPCR) superfamily (Figure 1). Like other GPCRs, these VNO receptors (VRs) have 7 hydrophobic stretches that may serve as membrane spanning domains. Only 287 of 850 residues are identical in all of the molecules shown in Figure1, indicating that the family is diverse. The VRs are related to two other types of GPCR, the calcium sensing receptor (CSR) and the metabotropic glutamate receptors (mGluRs) (Tanabe, Y., et al., *Neuron*, 1992, 8:169-179; Brown, E., et al., *Nature*, 1993, 366:575-580). The most highly related molecule is the CSR; for example, VR1 is 31% identical to rat CSR (Riccardi et al., *Proc. Natl. Acad. Sci. USA*, 1995, 92:131-135), with the highest homology residing in the TM1-TM7 region (44%) (Figure 1). However, the VRs comprise a distinct family of receptors, which share novel sequence motifs, and are more related to one another than they are to other receptors. For example, two divergent VRs, VR1 (SEQ ID NO. 1, 2) and VR4 (SEQ ID NO. 7, 8), are 70% identical in TM1-TM7, and 48% identical overall.

The VRs are unusual among GPCRs in having an extremely long N-terminal extracellular domain (Figures 1 and 2). This feature is shared by the CSR and mGluRs, and by an unrelated class of GPCRs that includes several receptors for glycoprotein hormones (Segaloff, D., et al., *Oxf. Rev. Reprod. Biol.*, 1992, 14:141-168). Importantly, the VRs are very different from both ORs and VNRs, which are also GPCRs (Buck. L., et al., *Cell*, 1991 51:127-133; Dulac, C., et al., *Cell*, 1995, 83:195-206). VRs share none of the characteristic sequence motifs of ORs or VNRs. In addition, the size of the N-terminal extracellular domain of VRs (557-565 amino acids) far exceeds that of ORs and VNRs (~12-28 amino acids) (Figure 2). The VRs are most variable in the N-terminal domain (25% identical residues compared to 57% in TM1-TM7). In the structurally-related mGluRs, the ligand binding site is thought to reside in the large N-terminal domain (O'Hara et al., *Neuron*, 1993, 11:41-52; Takahashi et al, *J. Biol. Chem.*, 1993, 268:19341-19345). If this is also true of VRs, the accentuated diversity of the N-terminal domain may reflect an ability to recognize diverse pheromonal ligands.

Most of the VR cDNAs that we analyzed appeared to belong to one of three subfamilies of highly related molecules. For example, VR1 (SEQ ID NOs. 1, 2), VR2 (SEQ ID NOs. 3, 4), and VR3 (SEQ ID NOs. 5, 6) are very similar as are VR4 (SEQ ID NOs. 7, 8) and VR5 (SEQ ID NOs. 9, 10), and VR6 (SEQ ID NOs. 11, 12) and VR7 (SEQ ID NOs. 13, 14) (Figure 1). Nonetheless, our results indicate that all of these cDNAs were derived from different genes. First, all cDNAs were sequenced on both strands to rule out sequencing errors. Second, the RNA used for library construction and PCR came from an inbred mouse strain (C57BL/6J), so they cannot be allelic variants. Third, the error rates of reverse transcriptase (or Taq polymerase) cannot account for the extent to which the cDNAs differ. For example,VR4 (SEQ ID NOs. 7, 8) and VR5 (SEQ ID NOs. 9, 10) cDNAs are 99% identical in nucleotide sequence, but the reverse transcriptase used to prepare them has an error rate of only 3.6×10^{-5}/bp (Ji, J., et al., *Biochemistry*,1992, 31:954-958).

Variant forms of VR mRNA

Many of the VRs we characterized lacked a segment of the N-terminal domain present in other VRs. Invariably, the missing segment corresponded to a region of the human CSR encoded by a single exon, or pair of exons (Pollak, M., et al., *Cell*, 1993, 73:1297-1303). We also found several different VR cDNAs that contained a stretch of noncoding sequence at a site corresponding to a CSR exon-intron boundary (e.g. VR15). This suggested that the exon-intron

structure of VR genes resembles that of the CSR gene, and that variant forms of VR mRNAs might be generated by differential RNA splicing.

Variant VR mRNAs could derive either from different genes, or from the same gene by alternative RNA splicing. Consistent with the latter possibility, two pairs of cDNAs that we sequenced VR8 (SEQ ID NOs. 15, 16) and VR9 (SEQ ID NOs. 17, 18), and VR10 (SEQ ID NOs. 19, 20) and VR11 (SEQ ID NOs. 21, 22) were identical in nucleotide sequence, but were missing different segments. However, when we used RT-PCR to amplify VNO mRNA sequences encoding 5 different VRs, we obtained one major PCR product in each case, regardless of whether the RNA used was from male or female mice. In 4 cases, the size of the major product corresponded to a complete VR, even though one of the cDNAs (but not the PCR product) contained an intron (#5). In one case, in which the cDNA lacked one exon (#2), the major PCR product was even smaller, and was found to lack two exons. Although PCR products of a smaller size were also seen in these experiments, they were much less abundant.

These results suggest that different VR forms derive from different genes. Thus many VR genes may be expressed pseudogenes, which either lack one or more exons, or have mutations that prevent proper RNA splicing. We cannot exclude the possibility that some variant VRs are functional, however. For example, some truncated VRs that lack transmembrane domains could conceivably be secreted pheromone-binding proteins.

Differential expression of VR genes in VNO neurons

To investigate the tissue distribution of VR gene expression, we conducted Northern blot analyses in which size fractionated polyA$^+$ RNAs from different mouse tissues were hybridized to a mix of radiolabeled VR cDNAs. The mixed probe hybridized to VNO RNAs of ~1.9-3.7 kb, with intense hybridization to RNAs of 2.8-3.5 kb. It did not hybridize to RNAs from a variety of other tissues, including olfactory epithelium and brain. This suggested that VR genes may be expressed exclusively in the VNO.

We found two partial cDNAs that were highly related to VR cDNAs in the NCBI dbEST database, one from spleen and the other from 2-cell stage mouse embryos. However, when we hybridized the most highly related VR cDNAs (VR6 and VR7) to spleen sections, only one questionably-labeled cell was seen out of ~1.4×10^6 cells with one VR probe, and none was seen with the other. The EST clones might be DNA contaminants, or be due to the widespread, but low level, misexpression of tissue specific genes (Sarkar, G., et al., *Science*, 1989, 244:331-334);

nonetheless, we cannot exclude the possibility that VR genes are expressed at a low frequency in some other tissues.

To examine the patterns of expression of different VR genes in the VNO, we conducted in situ hybridization experiments. Labeled segments of the 3' untranslated regions of three VR cDNAs were hybridized separately, or in combination, to sequential sections through the VNO. Probes prepared from $G_{\alpha o}$ and $G_{\alpha i2}$ cDNAs were hybridized to adjacent sections to delineate the $G_{\alpha o}$+ and $G_{\alpha i2}$+ zones of the VNO neuroepithelium.

The $G_{\alpha o}$ and $G_{\alpha i2}$ probes gave patterns of hybridization similar to those we had previously seen (Berghard, A., et al, *J. Neurosci.*, 1996, 16:909-918). The $G_{\alpha o}$ probe hybridized to a wavy stripe of VNO neurons in the basal (lower) region of the VNO neuroepithleium, whereas the $G_{\alpha i2}$ probe hybridized to an adjacent stripe of neurons in the apical (upper) part of the neuroepithelium. The waviness of the two zones appears to be caused by the periodic presence of blood vessels near the base of the epithelium (Berghard, A., et al, *J. Neurosci.*, 1996, 16:909-918). Approximately 57% of VNs were labeled by the $G_{\alpha i2}$ probe and 43% were labeled by the $G_{\alpha o}$ probe. The single layer of supporting cells located just beneath the epithelial surface was not labeled by either probe.

Each of the VR probes hybridized to a small percentage (2.4-5.7%) of VNs that appeared to be restricted to the basal, $G_{\alpha o}$+ zone of the VNO neuroepithelium. Labeled neurons were scattered throughout the anterior-posterior and dorsal-ventral extent of the $G_{\alpha o}$+ zone. Small clusters of labeled cells were somtimes seen, particularly with the VR2 probe The mixed probe labeled a larger percentage of VNs (10.6%) that was almost equal to the sum of the percentages labeled by its individual components (10.8%). Thus different $G_{\alpha o}$+ neurons must express different VRs.

No differences were seen in the patterns of hybridization obtained using VNOs from male and female mice, and no hybridization was observed in the nasal olfactory epithelium using either the mix of VR probes or a full-length VR cDNA probe (not shown). Subsequent analyses of the size of the VR gene family, and the number of VR genes recognized by the VR in situ hybridization probes, allowed us to estimate the number of VR genes expressed by individual neurons (see below).

The size f the VR multigen family

To investigate the size of the VR gene family, we hybridized several different mixed VR gene probes to a mouse genomic library, using high (70°C) or low (55°C) stringency conditions. A probe prepared from the membrane spanning regions (putative exon 6) of several different cDNA clones hybridized to 59 and 98 clones per haploid genome equivalent, at high and low stringency, respectively. To obtain probes that were potentially more diverse, we amplified internal segments of putative exon3 or 6 from genomic DNA by PCR with degenerate primers. At high stringency, these probes hybridized to 60-140 clones per haploid equivalent. These results indicate that there are as many as 140 VR genes in the mouse genome.

The VR probes that we used for in situ hybridization each labeled a small percentage of neurons. To determine how many VR genes each probe recognized, we hybridized probes prepared from the same VR cDNA segments to Southern blots of C57BL/6J mouse genomic DNA which had been digested with Eco RI or Hind III. Each probe hybridized to a small number of restriction fragments. Given the small size of the probes (~350-450 bp), most of these fragments should represent at least one gene, provided that there are no introns in the region probed. Consistent with this assumption, the VR2 (SEQ ID NO. 3) probe hybridized to 7 different restriction fragments, as many as five of which could be accounted for by characterized VR cDNAs that were 91-98% identical to VR2 (SEQ ID NO. 3) in the region probed.

Given the number of genes recognized by each VR probe and the percentage $G_{\alpha o}$+ neurons that hybridized to each, we estimate that each VR gene may be expressed in only ~1.1-1.9% of $G_{\alpha o}$+ VNs. Since there appear to be 60-140 VR genes in the mouse genome, this suggests that each $G_{\alpha o}$+ VNO neuron may express only one, or at most a few, VR genes.

Linkage of chromosomal clusters of VR and OR genes

We previously found that there are clusters of OR genes at multiple chromosomal sites in the mouse genome (Sullivan, S., et al., *Proc. Natl. Acad. Sci.*, 1996, 93:884-888). To investigate the chromosomal locations of VR genes, we used the Jackson Laboratory Backcross DNA Mapping Panel, which allows the mapping of mouse genes using interspecies mouse crosses.

Probes prepared from the 3' untranslated regions of VR2 (SEQ ID NO. 3) or VR4 cDNAs were first hybridized to Southern blots of genomic DNAs from two mouse species, C57BL/6J and Mus spretus, which had been digested with different restriction enzymes. Eco RI digests showed a number of restriction length polymorphisms with both VR probes. The VR probes

were then hybridized to Eco RI-digested DNAs from a large panel of different backcross mice ((C57BL/6J x M. spretus) x M. spretus).

The patterns of inheritance of the polymorphic fragments recognized by the two VR probes allowed us to assign chromosomal locations to approximately 9 VR genes. Using the VR4 (SEQ ID NO. 7) probe, we could follow the inheritance of 4 polymorphic restriction fragments. All of these cosegregated in the backcrosses, and mapped to the proximal end of chromosome 7 (near *D7Bir5)*. Five restriction fragments were followed for the VR2 (SEQ ID NO. 3) probe. Again, all of the restriction fragments cosegregated, allowing us to map the VR2 (SEQ ID NO. 3) fragments to the distal end of chromosome 4 (near *D4Bir1*). Given the resolution of the genetic mapping, the cosegregating fragments can be no more than 3.8 cM from one another. These results indicate that VR genes are located near the ends of at least two different mouse chromosomes. They also indicate that highly related VR genes are clustered at the same chromosomal locus, as previously seen in our studies and others (Ben-Arie et al, *Human Molecular Genetics,* 1994, 3:229-235.).

The VR4 gene subfamily appears to be closely linked to one OR gene locus, (*olfR5)* (Sullivan, S., et al., *Proc. Natl. Acad. Sci.*, 1996, 93:884-888). Although the VRs and ORs were mapped in different mouse crosses, the synaptotagmin-3 gene (*Syt3*) was mapped in both crosses, allowing an estimate of their relative positions. The OR locus mapped 15.05 cM proximal to *Syt3* while the VR4 gene cluster mapped 14.89 cM proximal to *Syt3*. (Jackson Laboratory Mouse Genome Informatics), suggesting a close linkage between VR and OR genes at the proximal end of chromosome 7. Our previous studies indicate that multiple OR gene loci arose via a series of duplications of very large chromosomal domains that maintained linkages between OR genes and members of other gene families. These results therefore suggest that VR genes and OR genes might have been linked in a primitive ancestor. They also suggest the possibility that additional clusters of VR genes might be linked to other OR gene loci.

Example 2

Experimental procedures

Preparation of cDNA Libraries from Isolated VNO Neurons

VNOs were dissected from adult (7- to 8-week-old) male Lewis rats (Sprague-Dawley). Single-cell cDNA synthesis and amplification were performed and checked according to Dulac and Axel (*Cell*,1995, 83:195-206). Southern blot analysis of single-cell cDNA was used to

detect expression of tubulin, OMP, Go, and $Gi_{2\alpha}$ (Dulac and Axel, *Cell*, 1995, 83:195-206). Eighteen cDNAs showed strong hybridization with tubulin and OMP probes, indicating that they originated from mature neurons, and were selected for further study. Cells VN3 and VN13 exhibited high levels of Go expression, whereas VN10 showed presence of $Gi_{2\alpha}$, indicating the origin of these cells from two distinct regions of the VNO neuroepithelium. VN13 single-cell cDNA library was prepared according to Dulac and Axel (*Cell*, 1995, 83:195-206).

Differential Screening of Single-Cell Library

Plaque-forming units (12×10^3) from the VN13 library were plated at low density, and duplicate filters (Hybond N^+, Amersham) were hybridized with probes generated from VN10 and VN13 single-cell cDNAs, following the procedure described in Dulac and Axel, *Cell*, 1995, 83:195-206. Ten phage plaques were detected that showed a positive signal unique to the VN13 probe. These plaques were purified, and the corresponding phage inserts were amplified by PCR, run on 1.5% agarose gel, blotted onto nylon filter, and hybridized with the VN10, VN3, and VN13 single-cell cDNA probes.

Isolation and Analysis of Full-Length cDNA Clones

A 425 bp clone, Go-VN13A, present at the frequency of 0.1% in the VN13 single-cell cDNA library, was selected and in vivo excised to generate the pBlueScriptSK(-) phagemid. High stringency (65°C) screening of a cDNA library prepared from female rat VNO (Dulac and Axel, *Cell*, 1995, 83:195-206) with the Go-VN13A cDNA probe led to the isolation of Go-VN13B (SEQ ID NO. 49) , presenting 90% sequence homology with Go-VN13A. Phages (7.2×10^5) of the female rat VNO library were further screened with the Go-VN13B (SEQ ID NO. 49) cDNA probe under low stringency conditions: hybridization was carried out at 55°C for 24 hr, and the filters were washed three times at 55°C for 30 min in 0.5x SSC and 0.5% SDS. A total of 75 positive phages were identified and the corresponding inserts were amplified by PCR and analyzed by Southern blot using the Go-VN13B (SEQ ID NO. 49) probe at both high (65°C) and low (55°C) stringency. This led to the identification of 22 cDNA clones with insert sizes longer than 3 kb. Among those, six distinct subfamilies were defined by absence of cross-hybridization under stringent conditions of hybridization and washing. Full-length clones (Go-VN1 to Go-VN6, SEQ ID NOs. 33, 35, 37, 39, 41, 43), each representative of a subfamily, were selected for *in vivo* excision and sequenced. Go-VN13C (SEQ ID NO. 47) and Go-VN13B

(SEQ ID NO. 49) are identical sequences differing by a 150 bp deletion in Go-VN13C (SEQ ID NO. 47). This sequence encodes for NMDQCANCPEYQYANTEKNKCIQKGVIVLSYEDPLGMALALIAFCFSAFTV (SEQ ID NO. 58) in Go-VN13B (SEQ ID NO. 49) and is replaced by an M at position 552 in Go-VN13C (SEQ ID NO. 48).

DNA Sequencing and Sequence Analysis

DNA sequencing was performed using ABI Prism dye terminator cycle ready reaction (Perkin Elmer, Norwalk, CT) according to manufacturer's protocol. Samples were run on an ABI Prism 310 Genetic Analyzer (Perkin Elmer, Norwalk, CT). Sequence homologies were determined using the BLAST system (NIH network service). Pairwise and ClustalW alignments (BLOSUM30 matrix setting) as well as Kyte-Doolittle hydropathic analysis were obtained with the MacVector sequence analysis software (Oxford Molecular Group).

In Situ Hybridization Analysis

In situ hybridization was performed as described elsewhere (Schaeren-Wiemers, N., et al., *Histochemistry*, 1993, 100:431-440). VNOs were dissected from adult male (8- to 9-week-old), adult female (9- to 11-week-old), and young (1-week-old) rats. Tissues were embedded in Tissue-Tek OCT. Antisense and sense digoxigenin-labeled probes were generated from the full-length cDNAs encoding for Go, $Gi_{2\alpha}$, Go-VN13B (SEQ ID NO. 49), and Go-VN1 to Go-VN6 (SEQ ID NOs. 33, 35, 37, 39, 41, 43), as well as from the 3' untranslated regions of the Go-VN1 to Go-VN6 clones.

Imaging Processing and Statistical Analysis

Digital photographs were captured with a Leitz DMRB microscope (Leica) coupled to a ProgRes3012 digital camera (Kontron Electronic) and further processed with the Photoshop (Adobe System) and Canvas (Deneba) software for Macintosh. The relative positions of cells exhibiting a positive signal by in situ hybridization were measured along the basal-apical axis using the NIH Image analysis software. The number of cells in hemiconcentric sections of 10% along this axis from the basal (value = 0) to the apical (value = 100) boundaries was determined. Average data for Go-VN1 and Go-VN3 to Go-VN6 were obtained from six to eight VNO sections, corresponding to four individuals analyzed in two independent experiments. For

Go-VN2, 14 VNO sections, corresponding to ten individuals and four independent experiments, were analyzed for each sex.

Southern Blot Analysis of Rat Genomic DNA and Screening of Rat and Human Genomic Libraries

Genomic DNA, prepared from Lewis rat (Sprague-Dawley) liver, was digested with the restriction enzymes EcoRI and BamHI, size fractionated on 0.8% agarose gels, and blotted onto nylon membrane (Sambrook, J., et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989). Membranes were cross-linked under UV light, hybridized overnight at both high (68°C) and low (55°C) stringency in hybridization buffer, and washed as described above. ^{32}P-labeled probes were generated by random priming, using the following DNA templates: EcoRI-EcoRV, NotI-NsiI, EcoRI-SalI, PstI-NdeI, XbaI-HincII, and EcoRI-NsiI fragments of Go-VN1 to Go-VN6 (SEQ ID NOs. 33, 35, 37, 39, 41, 43), respectively; a full-length (425 bp) insert of Go-VN13A; and a cDNA fragment including the seven transmembrane domains of Go-VN13B (SEQ ID NO. 49). Plaque-forming units (3 x 10^5) from rat and human genomic libraries (Stratagene, La Jolla, CA) were screened at low stringency (55°C) using a mix of ^{32}P-labeled probes prepared from fragments of Go-VN1 to Go-VN6 (SEQ ID NOs. 33, 35, 37, 39, 41, 43) encompassing the transmembrane domains 2 to 7.

Results

The VNO Neuroepithelium Expresses Two Independent Families of Pheromone Receptors

We hypothesized the existence of two distinct families of genes encoding pheromone receptor genes that are selectively colocalized with either the Go protein in the basal half of the vomeronasal neuroepithelium or with the $Gi_{2\alpha}$ protein in the apical region. For simplicity of nomenclature, and with the understanding that the cosegregation of distinct G-protein subunits with independent families of pheromone receptors is consistent but does not demonstrate a functional link, the family of genes encoding putative pheromone receptors that we have previously identified and that colocalize with $Gi_{2\alpha}$ will be named $Gi_{2\alpha}$-VN, whereas the novel family of receptors coexpressed with Go and described in this study will be named Go-VN. In the absence of information concerning the nature of the Go-VN receptor molecules, we reiterated the cloning strategy that allowed us to identify a family of putative pheromone receptor genes

expressed by $Gi_{2\alpha}$+ neurons (Dulac and Axel, *Cell*, 1995, 83:195-206). This strategy was based on the assumption that individual neurons within the VNO are likely to express only one pheromone receptor gene and that transcripts encoding a given receptor represent between 1% and 0.1% of a single-cell mRNA. Differential screening of cDNA libraries constructed from single-VNO neurons takes advantage of the fact that different cells express different receptors and thus provides an experimental solution to the problem of detecting a specific transcript in a heterogeneous population of neurons. In this attempt, we expected that differential screening of a cDNA library prepared from an isolated Go+, $Gi_{2\alpha}$- VNO neuron would permit the isolation of a class of pheromone receptor genes distinct from the $Gi_{2\alpha}$-VN family of receptor genes.

A cDNA library prepared from a Go+ neuron (VN13) was differentially hybridized with ^{32}P-labeled probes prepared from VN13 and from a second VNO neuron cDNA (VN10). A 425 bp cDNA (Go-VN13A) present at a frequency of 0.1% in the VN13-cDNA library showed selective hybridization with VN13 cell probe. Two cDNAs of longer size, Go-VN13B (SEQ ID NO. 49) and Go-VN13C (SEQ ID NO. 47), were subsequently isolated from a cDNA library prepared from dissected adult VNOs and showed 90% sequence similarity with Go-VN13A. Hybridization to VNO cross-sections with digoxigenin-labeled antisense RNA probe showed that expression of these transcripts is restricted to a small subpopulation of VNO neurons in a location consistent with the region of Go expression of the neuroepithelium. The sequence of Go-VN13B (SEQ ID NO. 49) reveals a partial open reading frame that includes seven hydrophobic stretches of 20 amino acids in length. Go-VN13B (SEQ ID NO. 49) sequence does not share any resemblance with the odorant receptor genes nor with the family of putative pheromone receptor genes previously identified (see below). In addition, hybridization of Go-VN13B DNA probe to genomic DNA identified two discrete bands at high stringency and 13 or more at lower stringency, revealing the existence of a family of closely related genes in the rat genome.

Taken together, these data indicate that we have isolated a novel multigene family encoding seven transmembrane domain receptors and expressed by subsets of VNO neurons from the basal half of the neuroepithelium.

Sequences f a New Family of VNO Receptors

Recombinant phages from a VNO cDNA library were screened at low stringency with the Go-VN13B (SEQ ID NO. 49) DNA probe. Six distinct gene subfamilies were isolated that

showed no cross-hybridization under stringent conditions of hybridization and washing. cDNAs Go-VN1 to Go-VN6, each representative of a subfamily, were fully sequenced (SEQ ID Nos 33, 35, 37, 39, 41 and 43).

In Go-VN1 to Go-VN5 cDNAs (SEQ ID Nos 33, 35, 37, 39 and 41), the first methionine of the open reading frame was tentatively chosen as a start for protein translation, revealing large open reading frames ranging from 548 to 866 amino acids. A frame shift in the Go-VN6 (SEQ ID NO. 44) sequence (amino acid 532; indicated by slash bar in Fig. 3) indicated that this transcript is unable to generate a functional protein.

Deduced Amino Acid Sequences of cDNAs from the Go-VN Family of Pheromone Receptors

The deduced amino acid sequences of eight cDNAs belonging to the Go-VN family of putative pheromone receptors is shown in Figure 3. Predicted position of seven transmembrane domains is also indicated (I-VII). Amino acids common to at least five cDNAs are shaded. Amino acids common to the rat mGluR1 and $Ca2^+$-sensing receptors are indicated by a star.

Hydropathy analysis of the predicted Go-VN proteins with the Kyte-Doolittle algorithm identified a large hydrophilic N-terminal domain that ranges in size from 274 amino acids in Go-VN1 (SEQ ID NO. 34) to 595 in Go-VN4 (SEQ ID NO. 40). This is preceded in cDNAs Go-VN4 (SEQ ID NO. 40), Go-VN7 (SEQ ID NO. 46), and Go-VN13C (SEQ ID NO. 50) by an initial hydrophobic 21 amino acid segment characteristic of eukaryotic signal sequences. A cluster of seven hydrophobic regions representing potential membrane-spanning helices and typical of the G protein-coupled receptor superfamily is followed by a short hydrophilic sequence that indicates a potential intracytoplasmic C-terminal domain. A database search indicated the presence of sequence motifs common to $Ca2^+$-sensing and metabotropic glutamate (mGluR) receptors (Houamed, K., et al., *Science*, 1991, 252:1318-1321; Masu, M., et al., *Nature*, 1991, 349:760-765; Brown, E., et al., *Nature*, 1993, 366:575-580 ; Pollak, M., et al., *Cell*, 1993 75:1297-1303). Pairwise sequence alignments reveal 18% to 23% sequence identity between the rat $Ca2^+$-sensing receptor and the most distant (Go-VN3, SEQ ID Nos.37, 38) and the closest (Go-VN1, SEQ ID NOs. 33, 34) Go-VN sequences, respectively. Sequences of rat mGluR1 and Go-VN cDNAs appear more distantly related. Several localized regions showed a more pronounced degree of similarity, including a cysteine-rich sequence just preceding the first transmembrane domain (amino acid 206 to 260 in Go-VN1, SEQ ID NO. 34), the predicted

transmembrane domains 2 to 7 with surrounding cytoplasmic and extracellular loops, and the relative position of 20 cysteines. The N-terminal and first transmembrane domains show little degree of homology. In mGluR and $Ca2^+$-sensing receptors, the second intracellular loop is involved in providing specificity for G-protein coupling (Gomeza, J., et al., *J. Biol. Chem.*, 1996, 271:2199-2205), enabling different classes of mGluR receptors to activate phospholipase C or to inhibit adenylyl cyclase. In Go-VN, this domain is rich in basic residues, as expected for potential G-protein coupling, and shows closer resemblance to the class II and III mGluRs that were shown to couple to Go and Gi subunits. Overall, the six Go-VN sequences share between 42% and 75% sequence identity. Regions of Go-VN proteins downstream of transmembrane domain 2 are nearly identical in all VNO receptor sequences. In contrast, N-terminal extracellular regions and first transmembrane domains are quite divergent.

Anomalies in Go-VN cDNA Sequences: Two unusual features were observed in the sequence of some Go-VN cDNAs. In Go-VN1 (SEQ ID NO. 33) and Go-VN3 (SEQ ID NO. 37) cDNAs, stretches of open reading frame can be found in the 5' extremity of the cDNAs that generate polypeptide sequences of 310 and and 152 amino acids, respectively, which are interrupted by a frameshift in Go-VN1 and by an insertion of 500 nucleic acids in Go-VN3. The prospective receptor protein sequences indicated for Go-VN1 (SEQ ID NO. 33) and Go-VN3 (SEQ ID NO. 37) (Fig. 3) start at the next available methionin and are therefore significantly shorter than those of other receptor cDNAs.

Go-VN7 (SEQ ID NO. 45) and Go-VN13C (SEQ ID NO. 47) cDNAs show a similar deletion of 150 bp located at the exact same position in the sequence. Strikingly, the 150 bp deletion does not alter the open reading frame but generates a gap that encompasses 34 amino acids upstream of the first transmembrane domain and most of the first transmembrane domain itself.

Hydropathy analysis of Go-VN7 (SEQ ID NO. 46) and Go-VN13C (SEQ ID NO. 48) protein sequences detects only a seven to eight amino acid long hydrophobic stretch that might not be long enough to replace the deleted transmembrane domain 1 and allow the appropriate folding of the protein. Except for the 150 bp gap, sequences of Go-VN13B (SEQ ID NO. 50) and Go-VN13C (SEQ ID NO. 48) are identical. This raises the question as to whether both transcripts might originate from alternative splicing of the same gene. Alternatively, they might be transcribed from independent genes that evolved from recent duplication and deletion events.

Size f the Go-VN Family of Genes

We investigated the size of the Go-VN family of receptors by hybridizing ^{32}P-labeled cDNA probes prepared from regions spanning the most divergent N-terminal half of the receptor protein to rat genomic DNA. Individual probes identify two to four discrete bands under stringent conditions of hybridization and washing. Under conditions of reduced stringency, each of the individual probes now generates a unique pattern of 12 to 20 bands, providing a direct illustration of the existence of a very large family of related genes.

A direct estimate of the size of the Go-VN receptor gene family was obtained by low stringency screening of a rat genomic library. PCR amplification on genomic DNA had indicated that receptor genes are devoid of introns in the region encompassing transmembrane domains 2 to 7, enabling us to deduce directly the number of genes present in the rat genome. A mix of ^{32}P-labeled DNA probes prepared from the six Go-VN cDNA fragments identified 110 positive clones per haploid genome, indicating that the family of Go-VN receptors may consist of 100 genes.

Expression Pattern of Go-VN Receptors

The pattern of expression of the Go-VN receptor genes was examined by in situ hybridization with digoxigenin-labeled RNA antisense probes. No signal was observed after hybridizing the mix of Go-VN1 to Go-VN6 (SEQ ID NOs. 33, 35, 37, 39, 41 and 43) receptor probes to sections of muscle, testis, brain, or whole head. The adult olfactory epithelium was also consistently negative, although rare positive cells (one to three cells per section) were observed in the olfactory neuroepithelium of E19 rat embryo. In contrast, strong signals were observed when antisense receptor RNA probes were hybridized to VNO neuroepithelium. In adults, each one of the Go-VN probes detects small subsets of VNO sensory neurons. When hybridization and washing were performed at lower temperature, the number of faintly labeled neurons increased, revealing cross- hybridization to more distant receptor genes.

Under high stringency conditions, cDNA clones Go-VN1 to Go-VN6 label 1.9%, 3.6%, 6.1%, 0.4%, 3.5%, and 1.3% of the VNO sensory neurons, respectively. Under the same experimental conditions, the mix of all six Go-VN RNA probes labels 19% of the cells. This number is similar to the sum of labeled neurons detected with the six individual Go-VN probes (17%), indicating that probes representing the six receptor subfamilies recognize distinct populations of VNO sensory neurons. Spatial Distribution of Go-VN Receptor Transcripts

Positive neurons identified with each of the Go-VN probes were randomly distributed along the anteroposterior and dorso-ventral axis of the VNO neuroepithelium. Most RNA probes recognize cells that are preferentially localized in the most basal two-thirds of the neuroepithelium corresponding to the zone of Go expression. However, careful examination of adjacent cross-sections of vomeronasal neuroepithelium labeled with each of the Go-VN probes reveals a well-organized spatial distribution of receptor expression. Different receptors appear preferentially localized in radial zones that define a series of hemiconcentric rings of distinct diameters. This pattern is observed along the entire length of the VNO and is conserved in all animals analyzed. The Go-VN3 (SEQ ID NO. 37) probe, for example, recognizes a subset of neurons that are confined to the most basal third of the VNO neuroepithelium. In contrast, the Go-VN1 (SEQ ID NO. 33), Go-VN4 (SEQ ID NO. 39), and Go-VN5 (SEQ ID NO. 41) RNA probes identify cells restricted to a hemiconcentric zone immediately apical to the area of Go-VN3 expression, whereas Go-VN2 identifies cells apposed to the apical layer of supporting cells. Go-VN6 in turn is found only in sparse cells immediately apposed to the basal membrane. This is best seen in a statistical representation of Go-VN receptor localization collected from VNO sections and multiple animals that shows a striking conservation of these patterns. Thus, transcription of Go-VN cDNAs appears restricted to one of three circumscribed areas of the VNO neuroepithelium in a manner quite reminiscent of the odorant receptor gene expression in four zones of the MOE (Ressler, K., et al., *Cell*, 1993, 73:597-609 ; Vassar, R., et al., *Cell*, 1993, 74:309-318). Although Go-VN3 (SEQ ID NO. 37) and Go-VN6 (SEQ ID NO. 43) transcripts show a clear segregation in the most basal region of the VNO neuroepithelium, the sequence anomalies found in both transcripts leave the functionality of this area of the neuroepithelium as an open question.

Sexual Dimorphism in Receptor Distribution and Age-Related Changes

To identify potential sexual dimorphism in Go-VN receptor expression, we systematically hybridized each probe to sections originating from adult male and female rat VNOs. All receptors were equally distributed in males and females with the striking exception of Go-VN2 (SEQ ID NO. 35). In females, Go-VN2 appears expressed in a large and centrally located region comprising one-third of the neuroepithelium. In sharp contrast, the same probe recognizes in males a cohort of cells in the most apical side of the neuroepithelium, closely apposed to the VNO lumen, and most likely intermingled with $Gi_{2\alpha}$ VNO sensory neurons. Such a difference

in the Go-VN2 expression pattern in males and females might result from the expression of the same receptor gene in a different zone of the VNO epithelium or from a differential expression of two distinct but closely related genes of the Go-VN2 subfamily. In females, Go-VN2 generates a very intense hybridization signal to most positive neurons and a fainter staining on a second set of labeled cells. The population of faintly labeled cells was never detected in males, indicating the existence of a female-specific neuronal subpopulation expressing either a lower level of the Go-VN2 transcript or a female-specific receptor significantly different but still cross-hybridizing to the Go-VN2 probe. We followed the emergence of receptor expression and of the VNO zonal organization during development and postnatal stages preceding puberty. Go-VN receptor expression is first detected in the VNO of E14 embryos. No significant difference is observed in the onset of expression of $Gi_{2\alpha}$-VN and Go-VN classes of receptor genes. In agreement with data of Berghard and Buck, 1996 in mouse, segregation of $Gi_{2\alpha}$ and Go expression in the apical and basal areas of VNO neuroepithelium, respectively, is not apparent in the embryo and in 1-week-old animals. In contrast, $Gi_{2\alpha}^{+}$ cells appear randomly distributed in large clusters over the whole thickness of the neuroepithelium, intermingled with Go cells. At 4 weeks after birth, however, $Gi_{2\alpha}$ cells appear clearly localized in the apex of the epithelium. Similarly, in situ hybridization experiments with mixes of Go-VN and $Gi_{2\alpha}$-VN receptor probes on sections of the VNOs dissected from late embryos and 1-week-old animals show that the two cell populations are still intermingled at early postnatal stages. We observed that the zonal distribution of the two families of receptors slowly emerges during sexual maturation to reach the spatial distribution observed in adults. Preliminary data indicate that the sexual dimorphic expression pattern of Go-VN2 is undetectable at 6 weeks after birth. Thus, in contrast to the zones of olfactory receptor gene expression, which are already present in the olfactory epithelium at the earliest stages of receptor gene expression in the embryo (Sullivan, S., et al., *Neuron*, 1995, 15:779-789), the spatial organization of the VNO neuroepithelium as detected by G-protein and receptor gene expression emerges only in a late postnatal period and reaches its definitive pattern at sexual maturity.

Expression of Go-VN Receptors Is Restricted to Go+ VNO Neurons

The expression of some of the Go-VN receptors in neurons lining the VNO lumen in an area mainly occupied by $Gi_{2\alpha}$+ cells raises the obvious question as to whether the expression of this family of genes is strictly restricted to Go+ VNO neurons. Single-cell cDNA prepared from

23 individual VNO neurons was analyzed by Southern blots with probes representing the six divergent subfamilies of Go-VN receptors and was PCR amplified with degenerated primers based on conserved motifs between Go-VN receptor sequences. Both approaches confirmed that none of the 19 cell cDNAs prepared from $Gi_{2\alpha}^{+}$ neurons contained any sequence of the Go-VN receptor family. In contrast, all four cDNAs generated from $Gi_{2\alpha}$- cells contained a sequence related to the Go-VN receptors. PCR products generated with degenerated primers based on conserved motifs between Go-VN receptor sequences and obtained from the four Go+ cells were subcloned and sequenced. For each single-cell cDNA, the insert sequences from ten independent colonies were found to be identical. This set of data strongly suggests that Go-VN receptor genes are not expressed by $Gi_{2\alpha}^{+}$ neurons and constitutes preliminary evidence for the expression of only one Go-VN receptor gene per neuron.

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims. All references disclosed herein are incorporated by reference in their entirety.

A Sequence Listing is presented below and is followed by what is claimed.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: PRESIDENT AND FELLOWS OF HARVARD COLLEGE

(ii) TITLE OF THE INVENTION: NOVEL PHEROMONE RECEPTORS

(iii) NUMBER OF SEQUENCES: 92

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
(B) STREET: 600 Atlantic Avenue
(C) CITY: Boston
(D) STATE: MA
(E) COUNTRY: U.S.A.
(F) ZIP: 02210-2211

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 60/051,284
(B) FILING DATE: 30-JUN-1997

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Plumer, Elizabeth R.
(B) REGISTRATION NUMBER: 36,637
(C) REFERENCE/DOCKET NUMBER: H0498/7074

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 617-720-3500
(B) TELEFAX: 617-720-2441
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3080 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 57...2606
(D) OTHER INFORMATION: VR1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GTTTTTCTGC ATCAGAAACG GATTTCACAG CAGCTCCATC TCAGATCCTA GCAGAC ATG 59

```
                                                                  Met
                                                                   1

AAG CAG CTC TGC GCT TTC ACT ATT TCT TTG TTG TTT CTG AAG TTT TCT      107
Lys Gln Leu Cys Ala Phe Thr Ile Ser Leu Leu Phe Leu Lys Phe Ser
            5                  10                  15

CTC ATC CTG TGC TGT TTG ACT GAA CCA AGT TGC TTT TGG AGA ATA AGG      155
Leu Ile Leu Cys Cys Leu Thr Glu Pro Ser Cys Phe Trp Arg Ile Arg
        20                  25                  30

AAT AGT GAA GAT AGT GAT GGA GAT TTA CAA AGG GAA TGT CAT TTT TAC      203
Asn Ser Glu Asp Ser Asp Gly Asp Leu Gln Arg Glu Cys His Phe Tyr
    35                  40                  45

CTT TGG AAA ACT GAT GAA CCT ATT GAA GAT AGT TTT TAT AAT TAT GAT      251
Leu Trp Lys Thr Asp Glu Pro Ile Glu Asp Ser Phe Tyr Asn Tyr Asp
50                  55                  60                  65

TTA AGT TTT AGA ATT GCA GCA AGT GAA TAT GAG TTT CTT CTC GTA ATG      299
Leu Ser Phe Arg Ile Ala Ala Ser Glu Tyr Glu Phe Leu Leu Val Met
                70                  75                  80

TTT TTT GCT ATC GAT GAG ATC AAC AGG AAT CCT TAT CTT TTA CCC AAC      347
Phe Phe Ala Ile Asp Glu Ile Asn Arg Asn Pro Tyr Leu Leu Pro Asn
            85                  90                  95

ATA ACT TTG ATG TTC TCC TTC ATT GGT GGA AAC TGT CAG GAT TTA TTG      395
Ile Thr Leu Met Phe Ser Phe Ile Gly Gly Asn Cys Gln Asp Leu Leu
        100                 105                 110

AGA GTT ATG GAC CAA GCA TAT ACA CAA ATA AAT GGA CAT ATG AAT TTT      443
Arg Val Met Asp Gln Ala Tyr Thr Gln Ile Asn Gly His Met Asn Phe
    115                 120                 125

GTT AAT TAT TTC TGT TAT TTA GAT GAT TCA TGT GCC ATA GGT CTT ACA      491
Val Asn Tyr Phe Cys Tyr Leu Asp Asp Ser Cys Ala Ile Gly Leu Thr
130                 135                 140                 145

GGA CCA TCA TGG AAA ACT TCC TTA AAA CTG GCA ATG CAC TCT TCG ATG      539
Gly Pro Ser Trp Lys Thr Ser Leu Lys Leu Ala Met His Ser Ser Met
                150                 155                 160

CCA CTG GTT TTC TTT GGA CCA TTT AAT CCT AAC CTA CGC GAC CAT GAC      587
Pro Leu Val Phe Phe Gly Pro Phe Asn Pro Asn Leu Arg Asp His Asp
            165                 170                 175

CGG CTG CCC CAT GTC CAT CAG GTA GCC CCC AAG GAC ACA CAT TTG TCC      635
Arg Leu Pro His Val His Gln Val Ala Pro Lys Asp Thr His Leu Ser
        180                 185                 190

CAT GGC ATG GTC TCC TTG ATG TTT CAC TTT AGA TGG ACT TGG ATA GGA      683
His Gly Met Val Ser Leu Met Phe His Phe Arg Trp Thr Trp Ile Gly
    195                 200                 205

CTG GTC ATC TCA GAT GAT GAC CAG GGT ATT CAG TTT CTC TCA GAT TTA      731
Leu Val Ile Ser Asp Asp Asp Gln Gly Ile Gln Phe Leu Ser Asp Leu
210                 215                 220                 225

AGA GAA GAA AGC CAA AGG CAT GGG ATC TGT TTA GCT TTT GTT AAT ATG      779
Arg Glu Glu Ser Gln Arg His Gly Ile Cys Leu Ala Phe Val Asn Met
                230                 235                 240

ATC CCA GAA AAC ATG CAG ATA TAC ATG ACA AGG GCT ACA ATA TAT GAT      827
Ile Pro Glu Asn Met Gln Ile Tyr Met Thr Arg Ala Thr Ile Tyr Asp
```

```
                245                 250                 255

AAA CAC ATT ATG ACA TCT TCA GCA AAG GTT GTT ATC ATT TAT GGT GAA      875
Lys His Ile Met Thr Ser Ser Ala Lys Val Val Ile Ile Tyr Gly Glu
        260                 265                 270

ATG AAC TCT ACT CTA GAA GCA AGC TTT AGA AGA TGG GAA GAG TTA GGT      923
Met Asn Ser Thr Leu Glu Ala Ser Phe Arg Arg Trp Glu Glu Leu Gly
    275                 280                 285

GCT CGG AGA ATC TGG ATC ACA ACC TCA CAA TGG GAT GTC ATC ACA AAT      971
Ala Arg Arg Ile Trp Ile Thr Thr Ser Gln Trp Asp Val Ile Thr Asn
290                 295                 300                 305

AAA AAA GAC TTC ACC CTT AAT CTC TTC CAT GGG ATC ATC ACT TTT GAA     1019
Lys Lys Asp Phe Thr Leu Asn Leu Phe His Gly Ile Ile Thr Phe Glu
                310                 315                 320

CAT CAT AGA TTT GAG ATT CCT AAA TTA AAT AAA TTC ATG CAA ACA ATG     1067
His His Arg Phe Glu Ile Pro Lys Leu Asn Lys Phe Met Gln Thr Met
            325                 330                 335

AAC ACT GCC AAA TAC CCA GTA GAT ATT TCT CAT ACT ATA TTG GAG TGG     1115
Asn Thr Ala Lys Tyr Pro Val Asp Ile Ser His Thr Ile Leu Glu Trp
        340                 345                 350

AAT TAT TTT AAT TGT TCA ATA TCT AAG AAC AGC ATT AGA ATG CAT CAT     1163
Asn Tyr Phe Asn Cys Ser Ile Ser Lys Asn Ser Ile Arg Met His His
    355                 360                 365

ATT ACA TTC AAC AAC ACC TTG GAA TGG ACA TCA CTG CAC AAC TAT GAT     1211
Ile Thr Phe Asn Asn Thr Leu Glu Trp Thr Ser Leu His Asn Tyr Asp
370                 375                 380                 385

GTG GCG ATG AGT GAT GAA GGT TAC AAT TTG TAC AAT GCT GTT TAT GCT     1259
Val Ala Met Ser Asp Glu Gly Tyr Asn Leu Tyr Asn Ala Val Tyr Ala
                390                 395                 400

GTG GCC CAC ACC TAC CAT GAA TAC ATT TTT CAA CAA GTA GAG TCT CAG     1307
Val Ala His Thr Tyr His Glu Tyr Ile Phe Gln Gln Val Glu Ser Gln
            405                 410                 415

AAA AAG GCA AAA CCC AAA AGA TAT TTC ACT GCT TGT CAG CAG GTG TCT     1355
Lys Lys Ala Lys Pro Lys Arg Tyr Phe Thr Ala Cys Gln Gln Val Ser
        420                 425                 430

TCC TTG ATG AAA ACC AGG GTA TTT ACG AAC CCT GTT GGA GAA CTG GTG     1403
Ser Leu Met Lys Thr Arg Val Phe Thr Asn Pro Val Gly Glu Leu Val
    435                 440                 445

AAC ATG AAG CAT AGG GAA AAT CAG TGT ACA GAG TAT GAT ATT TTC ATC     1451
Asn Met Lys His Arg Glu Asn Gln Cys Thr Glu Tyr Asp Ile Phe Ile
450                 455                 460                 465

ATT TGG AAT TTT CCA CAA GGC CTT GGA TTA AAA GTG AAA ATA GGA AGC     1499
Ile Trp Asn Phe Pro Gln Gly Leu Gly Leu Lys Val Lys Ile Gly Ser
                470                 475                 480

TAT TTA CCT TGT TTT CCA CAG AGA CAA AAA CTT CAT ATA TCT GAT GAT     1547
Tyr Leu Pro Cys Phe Pro Gln Arg Gln Lys Leu His Ile Ser Asp Asp
            485                 490                 495

TTG GAA TGG GCC AAG GGA GGA ACA TCA CCT CAG GTT CCC TCC TCC GTG     1595
Leu Glu Trp Ala Lys Gly Gly Thr Ser Pro Gln Val Pro Ser Ser Val
        500                 505                 510
```

```
TGT AGT GTG GCA TGT ACT GCT GGA TTC AGG AAA ATT TAT CAA AAA GAA    1643
Cys Ser Val Ala Cys Thr Ala Gly Phe Arg Lys Ile Tyr Gln Lys Glu
    515                 520                 525

ACA GCA GAC TGC TGC TTT GAT TGT GTT CAG TGC CCA GAA AAT GAG ATT    1691
Thr Ala Asp Cys Cys Phe Asp Cys Val Gln Cys Pro Glu Asn Glu Ile
530                 535                 540                 545

TCC AAC GAA ACA GAT ATG GAA CAG TGT GTG AGG TGT CCA GAT GAT AAG    1739
Ser Asn Glu Thr Asp Met Glu Gln Cys Val Arg Cys Pro Asp Asp Lys
                550                 555                 560

TAT GCC AAC ATA GAG CAA ACC CAC TGC CTC TCA AGA GCT GTA TCA TTT    1787
Tyr Ala Asn Ile Glu Gln Thr His Cys Leu Ser Arg Ala Val Ser Phe
            565                 570                 575

CTG GCT TAT GAA GAT TCA TTG GGG ATG GCT CTA GGC TGC ATG GCA CTG    1835
Leu Ala Tyr Glu Asp Ser Leu Gly Met Ala Leu Gly Cys Met Ala Leu
        580                 585                 590

TCC TTC TCA GCC ATC ACA ATT CTA ATC CTC GTC ACA TTT GTG AAG TAC    1883
Ser Phe Ser Ala Ile Thr Ile Leu Ile Leu Val Thr Phe Val Lys Tyr
    595                 600                 605

AAA GAT ACT CCC ACT GTG AAG GCC AAT AAC CGC ATT CTC AGC TAC ATC    1931
Lys Asp Thr Pro Thr Val Lys Ala Asn Asn Arg Ile Leu Ser Tyr Ile
610                 615                 620                 625

CTG CTC ATC TCT CTC GTC TTC TGC TTT CTC TGC TCC CTG CTC TTC ATT    1979
Leu Leu Ile Ser Leu Val Phe Cys Phe Leu Cys Ser Leu Leu Phe Ile
                630                 635                 640

GGA CCT CCC GAC CAG GTC ACC TGC ATC TTT CAG CAG ACC ACA TTT GGA    2027
Gly Pro Pro Asp Gln Val Thr Cys Ile Phe Gln Gln Thr Thr Phe Gly
            645                 650                 655

GTA TTG TTC ACT GTG TCT GTT TCT ACA GTG TTG GCC AAA ACA ATA ACT    2075
Val Leu Phe Thr Val Ser Val Ser Thr Val Leu Ala Lys Thr Ile Thr
        660                 665                 670

GTG GTC ATG GCT TTC AAG CTC ACT ACT CCA GGA AGA AGG ATG AGA GGG    2123
Val Val Met Ala Phe Lys Leu Thr Thr Pro Gly Arg Arg Met Arg Gly
    675                 680                 685

ATG ATG ATG ACA GGG GCA CCT AAG TTG GTC ATT CCC ATT TGT ACC CTG    2171
Met Met Met Thr Gly Ala Pro Lys Leu Val Ile Pro Ile Cys Thr Leu
690                 695                 700                 705

ATC CAA CTT GTT CTC TGT GGA ATC TGG TTG GTC ACA TCT CCT CCC TTT    2219
Ile Gln Leu Val Leu Cys Gly Ile Trp Leu Val Thr Ser Pro Pro Phe
                710                 715                 720

ATT GAC AGA GAC ATA CAA TCT GAG CAT GGG AAG ATT GTC ATT CTT TGC    2267
Ile Asp Arg Asp Ile Gln Ser Glu His Gly Lys Ile Val Ile Leu Cys
            725                 730                 735

AAT AAA GGC TCA GTC ATT GCC TTC CAC GTC GTC CTG GGA TAC TTG GGC    2315
Asn Lys Gly Ser Val Ile Ala Phe His Val Val Leu Gly Tyr Leu Gly
        740                 745                 750

TCC TTG GCT CTG GGG AGC TTC ACG TTG GCT TTC CTG GCT AGG AAC CTT    2363
Ser Leu Ala Leu Gly Ser Phe Thr Leu Ala Phe Leu Ala Arg Asn Leu
    755                 760                 765

CCT GAC ACA TTC AAT GAA GCC AAG TTC CTA ACT TTC AGC ATG CTG GTG    2411
```

```
Pro Asp Thr Phe Asn Glu Ala Lys Phe Leu Thr Phe Ser Met Leu Val
770                 775                 780                 785

TTC TGC AGT GTC TGG ATC ACC TTC CTC CCT GTC TAC CAC AGC ACC AGG    2459
Phe Cys Ser Val Trp Ile Thr Phe Leu Pro Val Tyr His Ser Thr Arg
                790                 795                 800

GGG AGG GTC ATG GTG GTT GTG GAG GTT TTC TCC ATC TTG GCT TCT AGT    2507
Gly Arg Val Met Val Val Val Glu Val Phe Ser Ile Leu Ala Ser Ser
            805                 810                 815

GCA GGG TTG CTA ATG TGT ATC TTT GTC CCA AAG TGT TAT GTT ATT TTA    2555
Ala Gly Leu Leu Met Cys Ile Phe Val Pro Lys Cys Tyr Val Ile Leu
        820                 825                 830

ATT AGA CCA GAT TCA AAT TTT ATA AAG AAC CAC AAA GGT AAA TTG CTT    2603
Ile Arg Pro Asp Ser Asn Phe Ile Lys Asn His Lys Gly Lys Leu Leu
    835                 840                 845

TAT TGAAACTTTC ATGGTATGAA AATGTTAGAT GATATTCAAC TTATCTTATT CTTCAT  2662
Tyr
850

CTTAATAAAA GCAGTACTTC ATCATATAAA AAATAAAGTA ATATACAGAT TTATACTTAC   2722
AAACTGGACA GCAAACATGA ATATGTTGAG AACTGGGATT CTCAATTGAG GAATGGCTAC   2782
CAATATTTTG ATCTGTGGTT TTGTGTTTAA GCCATGTACT TAATTAATGA TTAATATGAG   2842
GTTACCCTAC TGTCTTTGAA CAGCGCCACC TCTAGGCATG CTGTCCTTGA GTTATAAGAA   2902
AGGGTACTGC ATACACAATG GACATGAAGC CAGTAATCAA CATTATTCCA CTTGCTTTCA   2962
TGGAGTTCTT ACATCCAAGT TCATGCCTTG ACTTTATTCA ATGTTCTATG ACAAAGGTAG   3022
ATAAATAAAT AAACACTTTC CTCGTCGACG CGGCCGCGTC GACGTCGACG CGGCCGCG     3080
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 850 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Lys Gln Leu Cys Ala Phe Thr Ile Ser Leu Leu Phe Leu Lys Phe
 1               5                  10                  15
Ser Leu Ile Leu Cys Cys Leu Thr Glu Pro Ser Cys Phe Trp Arg Ile
            20                  25                  30
Arg Asn Ser Glu Asp Ser Asp Gly Asp Leu Gln Arg Glu Cys His Phe
        35                  40                  45
Tyr Leu Trp Lys Thr Asp Glu Pro Ile Glu Asp Ser Phe Tyr Asn Tyr
    50                  55                  60
Asp Leu Ser Phe Arg Ile Ala Ala Ser Glu Tyr Glu Phe Leu Leu Val
65                  70                  75                  80
Met Phe Phe Ala Ile Asp Glu Ile Asn Arg Asn Pro Tyr Leu Leu Pro
                85                  90                  95
Asn Ile Thr Leu Met Phe Ser Phe Ile Gly Gly Asn Cys Gln Asp Leu
            100                 105                 110
Leu Arg Val Met Asp Gln Ala Tyr Thr Gln Ile Asn Gly His Met Asn
        115                 120                 125
Phe Val Asn Tyr Phe Cys Tyr Leu Asp Asp Ser Cys Ala Ile Gly Leu
    130                 135                 140
Thr Gly Pro Ser Trp Lys Thr Ser Leu Lys Leu Ala Met His Ser Ser
145                 150                 155                 160
Met Pro Leu Val Phe Phe Gly Pro Phe Asn Pro Asn Leu Arg Asp His
```

```
                    165                 170                 175
Asp Arg Leu Pro His Val His Gln Val Ala Pro Lys Asp Thr His Leu
            180                 185                 190
Ser His Gly Met Val Ser Leu Met Phe His Phe Arg Trp Thr Trp Ile
        195                 200                 205
Gly Leu Val Ile Ser Asp Asp Asp Gln Gly Ile Gln Phe Leu Ser Asp
    210                 215                 220
Leu Arg Glu Glu Ser Gln Arg His Gly Ile Cys Leu Ala Phe Val Asn
225                 230                 235                 240
Met Ile Pro Glu Asn Met Gln Ile Tyr Met Thr Arg Ala Thr Ile Tyr
                245                 250                 255
Asp Lys His Ile Met Thr Ser Ser Ala Lys Val Val Ile Ile Tyr Gly
            260                 265                 270
Glu Met Asn Ser Thr Leu Glu Ala Ser Phe Arg Arg Trp Glu Glu Leu
        275                 280                 285
Gly Ala Arg Arg Ile Trp Ile Thr Thr Ser Gln Trp Asp Val Ile Thr
    290                 295                 300
Asn Lys Lys Asp Phe Thr Leu Asn Leu Phe His Gly Ile Ile Thr Phe
305                 310                 315                 320
Glu His His Arg Phe Glu Ile Pro Lys Leu Asn Lys Phe Met Gln Thr
                325                 330                 335
Met Asn Thr Ala Lys Tyr Pro Val Asp Ile Ser His Thr Ile Leu Glu
            340                 345                 350
Trp Asn Tyr Phe Asn Cys Ser Ile Ser Lys Asn Ser Ile Arg Met His
        355                 360                 365
His Ile Thr Phe Asn Asn Thr Leu Glu Trp Thr Ser Leu His Asn Tyr
    370                 375                 380
Asp Val Ala Met Ser Asp Glu Gly Tyr Asn Leu Tyr Asn Ala Val Tyr
385                 390                 395                 400
Ala Val Ala His Thr Tyr His Glu Tyr Ile Phe Gln Gln Val Glu Ser
                405                 410                 415
Gln Lys Lys Ala Lys Pro Lys Arg Tyr Phe Thr Ala Cys Gln Gln Val
            420                 425                 430
Ser Ser Leu Met Lys Thr Arg Val Phe Thr Asn Pro Val Gly Glu Leu
        435                 440                 445
Val Asn Met Lys His Arg Glu Asn Gln Cys Thr Glu Tyr Asp Ile Phe
    450                 455                 460
Ile Ile Trp Asn Phe Pro Gln Gly Leu Gly Leu Lys Val Lys Ile Gly
465                 470                 475                 480
Ser Tyr Leu Pro Cys Phe Pro Gln Arg Gln Lys Leu His Ile Ser Asp
                485                 490                 495
Asp Leu Glu Trp Ala Lys Gly Gly Thr Ser Pro Gln Val Pro Ser Ser
            500                 505                 510
Val Cys Ser Val Ala Cys Thr Ala Gly Phe Arg Lys Ile Tyr Gln Lys
        515                 520                 525
Glu Thr Ala Asp Cys Cys Phe Asp Cys Val Gln Cys Pro Glu Asn Glu
    530                 535                 540
Ile Ser Asn Glu Thr Asp Met Glu Gln Cys Val Arg Cys Pro Asp Asp
545                 550                 555                 560
Lys Tyr Ala Asn Ile Glu Gln Thr His Cys Leu Ser Arg Ala Val Ser
                565                 570                 575
Phe Leu Ala Tyr Glu Asp Ser Leu Gly Met Ala Leu Gly Cys Met Ala
            580                 585                 590
Leu Ser Phe Ser Ala Ile Thr Ile Leu Ile Leu Val Thr Phe Val Lys
        595                 600                 605
Tyr Lys Asp Thr Pro Thr Val Lys Ala Asn Asn Arg Ile Leu Ser Tyr
    610                 615                 620
Ile Leu Leu Ile Ser Leu Val Phe Cys Phe Leu Cys Ser Leu Leu Phe
625                 630                 635                 640
Ile Gly Pro Pro Asp Gln Val Thr Cys Ile Phe Gln Gln Thr Thr Phe
                645                 650                 655
Gly Val Leu Phe Thr Val Ser Val Ser Thr Val Leu Ala Lys Thr Ile
            660                 665                 670
Thr Val Val Met Ala Phe Lys Leu Thr Thr Pro Gly Arg Arg Met Arg
        675                 680                 685
```

```
Gly Met Met Met Thr Gly Ala Pro Lys Leu Val Ile Pro Ile Cys Thr
    690                 695                 700
Leu Ile Gln Leu Val Leu Cys Gly Ile Trp Leu Val Thr Ser Pro Pro
705                 710                 715                 720
Phe Ile Asp Arg Asp Ile Gln Ser Glu His Gly Lys Ile Val Ile Leu
                725                 730                 735
Cys Asn Lys Gly Ser Val Ile Ala Phe His Val Val Leu Gly Tyr Leu
            740                 745                 750
Gly Ser Leu Ala Leu Gly Ser Phe Thr Leu Ala Phe Leu Ala Arg Asn
        755                 760                 765
Leu Pro Asp Thr Phe Asn Glu Ala Lys Phe Leu Thr Phe Ser Met Leu
    770                 775                 780
Val Phe Cys Ser Val Trp Ile Thr Phe Leu Pro Val Tyr His Ser Thr
785                 790                 795                 800
Arg Gly Arg Val Met Val Val Val Glu Val Phe Ser Ile Leu Ala Ser
                805                 810                 815
Ser Ala Gly Leu Leu Met Cys Ile Phe Val Pro Lys Cys Tyr Val Ile
            820                 825                 830
Leu Ile Arg Pro Asp Ser Asn Phe Ile Lys Asn His Lys Gly Lys Leu
        835                 840                 845
Leu Tyr
    850
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2961 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 86...2509
(D) OTHER INFORMATION: VR2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
AGACACATCG GTGCAACTGT GTGTGTGATG TTTTTCTGCA TCAGAAACGG ATTTCACAGC      60
AGCTCCATCT CAGATCCTAG CAGAC ATG AAG CAG CTC TGC ACT TTC ACT ATT      112
                            Met Lys Gln Leu Cys Thr Phe Thr Ile
                             1               5

TCA TTG TTG TTT CTG AAG TTT TCT CTC ATC TTG TGC TGT TGG AGT GAA      160
Ser Leu Leu Phe Leu Lys Phe Ser Leu Ile Leu Cys Cys Trp Ser Glu
10                  15                  20                  25

CCA AGC TGC TTT TGG AGG ATA AAG AAG AGT GAA GAT AAT GAT GGA GAT      208
Pro Ser Cys Phe Trp Arg Ile Lys Lys Ser Glu Asp Asn Asp Gly Asp
                30                  35                  40

TTA CAA AGG GAG TGT CAT TTT TAC CTT TGG AAA ACT GAT GAA CCT ATT      256
Leu Gln Arg Glu Cys His Phe Tyr Leu Trp Lys Thr Asp Glu Pro Ile
            45                  50                  55

GAA GAT AGT TTT TAT AAT TAT GAT TTA AGT TTT AGA ATT GCA GGA AGT      304
Glu Asp Ser Phe Tyr Asn Tyr Asp Leu Ser Phe Arg Ile Ala Gly Ser
        60                  65                  70

GAA TAT GAG CTT CTT CTG GTA ATG TTT TTT GCT ACT GAT GAG ATC AAC      352
Glu Tyr Glu Leu Leu Leu Val Met Phe Phe Ala Thr Asp Glu Ile Asn
    75                  80                  85
```

```
AAG AAT CCT TAT CTT TTA CCC AAC ATG AGT TTG ATG TTC TCC ATC ATT      400
Lys Asn Pro Tyr Leu Leu Pro Asn Met Ser Leu Met Phe Ser Ile Ile
90                  95                  100                 105

GGT GGA AAC TGT CAT GAT TTA TTG AGA AGT CTG GAT CAA GAA TAT GCA      448
Gly Gly Asn Cys His Asp Leu Leu Arg Ser Leu Asp Gln Glu Tyr Ala
                110                 115                 120

CAA ATA GAT GGA CAT ATG AAT TTT GTT AAT TAT TTC TGT TAT TTA GAT      496
Gln Ile Asp Gly His Met Asn Phe Val Asn Tyr Phe Cys Tyr Leu Asp
            125                 130                 135

GAT TCA TGT GCC ACA GGC CTT ACA GGA CCA TCA TGG AAA ACA TCC TTA      544
Asp Ser Cys Ala Thr Gly Leu Thr Gly Pro Ser Trp Lys Thr Ser Leu
        140                 145                 150

AAA CTG GCA ATG CAT TCT TCA ATG CCA CTG GTT TTC TTT GGA CCA TTT      592
Lys Leu Ala Met His Ser Ser Met Pro Leu Val Phe Phe Gly Pro Phe
    155                 160                 165

AAT CCT AAC CTA CGC GAC CAT GAC CGG CTG CCC CAT GTC CAT CAG GTA      640
Asn Pro Asn Leu Arg Asp His Asp Arg Leu Pro His Val His Gln Val
170                 175                 180                 185

GCC CCC AAG GAC ACA CAT TTG TCC CAT GGC ATG GTC TCC TTG ATG TTT      688
Ala Pro Lys Asp Thr His Leu Ser His Gly Met Val Ser Leu Met Phe
                190                 195                 200

CAT TTT AGG TGG ACT TGG ATA GGA CTG GTC ATC TCA GAT GAT GAT CAG      736
His Phe Arg Trp Thr Trp Ile Gly Leu Val Ile Ser Asp Asp Asp Gln
            205                 210                 215

GGT ATT CAG TTT CTC TCA GAT TTA AGA GAA GAA AGC CAA AGG CAT GGG      784
Gly Ile Gln Phe Leu Ser Asp Leu Arg Glu Glu Ser Gln Arg His Gly
        220                 225                 230

ATC TGT TTG GCT TTT GTT AAT ATG ATC CCA GAA AAC ATG CAG ATA TAC      832
Ile Cys Leu Ala Phe Val Asn Met Ile Pro Glu Asn Met Gln Ile Tyr
    235                 240                 245

ATG ACA AGG GCT ACA ATA TAT GAT ACA CAA ATT ATG ACA TCT TCA GCA      880
Met Thr Arg Ala Thr Ile Tyr Asp Thr Gln Ile Met Thr Ser Ser Ala
250                 255                 260                 265

AAG GTT GTT ATC ATT TAT GGT GAC ATG AAC TCT ACT CTA GAA GCA AGC      928
Lys Val Val Ile Ile Tyr Gly Asp Met Asn Ser Thr Leu Glu Ala Ser
                270                 275                 280

TTT AGA AGA TGG GAA GAG TTA GGT GCT CGG AGA ATC TGG ATC ACA ACC      976
Phe Arg Arg Trp Glu Glu Leu Gly Ala Arg Arg Ile Trp Ile Thr Thr
            285                 290                 295

ACA CAA TGG GAT GTC ATC ACA AAT AAA AAA GAC TTC ACC CTT AAT CTC     1024
Thr Gln Trp Asp Val Ile Thr Asn Lys Lys Asp Phe Thr Leu Asn Leu
        300                 305                 310

TTC CAT GGG ACT ATT ACT TTT GCA CAC CAC AAA GAT GAG ATT CCT AAA     1072
Phe His Gly Thr Ile Thr Phe Ala His His Lys Asp Glu Ile Pro Lys
    315                 320                 325

TTT AGG AAT TTT ATG CAA ACA AAG AAA ACT GCC AAA TAC CTT GTA GAT     1120
Phe Arg Asn Phe Met Gln Thr Lys Lys Thr Ala Lys Tyr Leu Val Asp
330                 335                 340                 345

ATT TCT CAT ACT ATT TTG GAG TGG AAT TAT TTT AAT TGT TCA ATC TCT     1168
```

```
Ile Ser His Thr Ile Leu Glu Trp Asn Tyr Phe Asn Cys Ser Ile Ser
                350                 355                 360

AAG AAC AGC AGT AAA ATG GGT CAT TTT ACA TTC AAC AAC ACA TTG CAA    1216
Lys Asn Ser Ser Lys Met Gly His Phe Thr Phe Asn Asn Thr Leu Gln
            365                 370                 375

TGG ACA GCA CTG CAC AAC TAT GAT ATG GCC CTG AGC GAT GAA GGT TAC    1264
Trp Thr Ala Leu His Asn Tyr Asp Met Ala Leu Ser Asp Glu Gly Tyr
        380                 385                 390

AAT TTG TAT AAT GCT GTT TAT GCT GTG GCC CAC ACC TAC CAT GAA TAC    1312
Asn Leu Tyr Asn Ala Val Tyr Ala Val Ala His Thr Tyr His Glu Tyr
    395                 400                 405

ATT CTT CAA CAA GTA GAG TCT CAG AAA AAG GCA AAA CCC AAA AGA TAT    1360
Ile Leu Gln Gln Val Glu Ser Gln Lys Lys Ala Lys Pro Lys Arg Tyr
410                 415                 420                 425

TTC ACT GCT TGT CAG CAG GTG TCT TCC TTG ATG AAA ACC AGG GTA TTT    1408
Phe Thr Ala Cys Gln Gln Val Ser Ser Leu Met Lys Thr Arg Val Phe
                430                 435                 440

ATG AAC CCT GTT GGA GAA CTG GTG AAC ATG AAG CAT AGG GAA AAT CAG    1456
Met Asn Pro Val Gly Glu Leu Val Asn Met Lys His Arg Glu Asn Gln
            445                 450                 455

TGT ACA GAG TAT GAT ATT TTC ATC ATT TGG AAT TTT CCA CAA GGC CTT    1504
Cys Thr Glu Tyr Asp Ile Phe Ile Ile Trp Asn Phe Pro Gln Gly Leu
        460                 465                 470

GGA TTA AAA GTG AAA GTA GGA AGC TAT TTA CCT TGC TTT CCA AAG AGT    1552
Gly Leu Lys Val Lys Val Gly Ser Tyr Leu Pro Cys Phe Pro Lys Ser
    475                 480                 485

CAA CAA CTT CAT ATA GCT GAT GAT TTG GAA TGG GCC ATG GGA GGA ACA    1600
Gln Gln Leu His Ile Ala Asp Asp Leu Glu Trp Ala Met Gly Gly Thr
490                 495                 500                 505

TCA GTG GAT ATG GAA CAG TGT GTG AGA TGT CCA GAT AAT AAA TAT GCC    1648
Ser Val Asp Met Glu Gln Cys Val Arg Cys Pro Asp Asn Lys Tyr Ala
                510                 515                 520

AAT TTA GAG CAA ACC CAC TGC CTC CAA AGA ACG GTG TCA TTT CTG GCT    1696
Asn Leu Glu Gln Thr His Cys Leu Gln Arg Thr Val Ser Phe Leu Ala
            525                 530                 535

TAT GAA GAT CCA TTG GGG ATG GCT CTA GGC TGC ATG GCA CTG TCC TTC    1744
Tyr Glu Asp Pro Leu Gly Met Ala Leu Gly Cys Met Ala Leu Ser Phe
        540                 545                 550

TCG GCC ATC ACA ATT CTA GTC CTC GTC ACA TTT GTG AAG TAC AAG GAT    1792
Ser Ala Ile Thr Ile Leu Val Leu Val Thr Phe Val Lys Tyr Lys Asp
    555                 560                 565

ACT CCC ATT GTG AAG GCC AAT AAC CGC ATT CTC AGC TAC ATC CTG CTC    1840
Thr Pro Ile Val Lys Ala Asn Asn Arg Ile Leu Ser Tyr Ile Leu Leu
570                 575                 580                 585

ATC TCT CTC GTC TTC TGC TTT CTC TGT TCC CTG CTC TTC ATT GGA CAT    1888
Ile Ser Leu Val Phe Cys Phe Leu Cys Ser Leu Leu Phe Ile Gly His
                590                 595                 600

CCC GAC CAG GTC ACC TGC ATC TTG CAG CAG ACC ACA TTT GGA GTA TTG    1936
Pro Asp Gln Val Thr Cys Ile Leu Gln Gln Thr Thr Phe Gly Val Leu
```

```
            605                 610                 615

TTC ACT GTG TCT GTT TCT ACA GTG TTG GCC AAA ACA ATA ACT GTG GTC     1984
Phe Thr Val Ser Val Ser Thr Val Leu Ala Lys Thr Ile Thr Val Val
        620                 625                 630

ATG GCT TTC AAG CTC ACT ACT CCA GGA AGA AGG ATG AGA GGG ATG ATG     2032
Met Ala Phe Lys Leu Thr Thr Pro Gly Arg Arg Met Arg Gly Met Met
    635                 640                 645

ATG ACA GGG GCA CCT AAG TTG GTC ATT CCC ATT TGT ACC CTG ATC CAA     2080
Met Thr Gly Ala Pro Lys Leu Val Ile Pro Ile Cys Thr Leu Ile Gln
650                 655                 660                 665

CTT GTT CTC TGT GGA ATC TGG TTG GTC ACA TCT CCT CCC TTT ATT GAC     2128
Leu Val Leu Cys Gly Ile Trp Leu Val Thr Ser Pro Pro Phe Ile Asp
                670                 675                 680

AGA GAT ATA CAA TCT GAA CAT GGG AAG ATT GTC ATT CTT TGC AAT AAA     2176
Arg Asp Ile Gln Ser Glu His Gly Lys Ile Val Ile Leu Cys Asn Lys
            685                 690                 695

GGC TCT GTC GTT GCC TTC CAC GTC GTC CTG GGA TAC TTG GGC TCC TTG     2224
Gly Ser Val Val Ala Phe His Val Val Leu Gly Tyr Leu Gly Ser Leu
        700                 705                 710

GCT CTG GGG AGC TTC ACT TTG GCT TTC TTG GCT AGG AAC CTT CCT GAC     2272
Ala Leu Gly Ser Phe Thr Leu Ala Phe Leu Ala Arg Asn Leu Pro Asp
    715                 720                 725

ACA TTC AAT GAA GCC AAG TTC CTA ACT TTC AGC ATG CTG GTG TTC TGC     2320
Thr Phe Asn Glu Ala Lys Phe Leu Thr Phe Ser Met Leu Val Phe Cys
730                 735                 740                 745

AGT GTC TGG ATC ACC TTC CTC CCT GTC TAC CAC AGC ACC AGG GGG AAG     2368
Ser Val Trp Ile Thr Phe Leu Pro Val Tyr His Ser Thr Arg Gly Lys
                750                 755                 760

GTC ATG GTG GTT GTG GAG GTT TTC TCC ATC TTG GCT TCT AGT GCA GGG     2416
Val Met Val Val Val Glu Val Phe Ser Ile Leu Ala Ser Ser Ala Gly
            765                 770                 775

TTG CTA ATG TGT ATC TTT GTC CCA AAG TGT TAT GTT ATT TTA ATT AGA     2464
Leu Leu Met Cys Ile Phe Val Pro Lys Cys Tyr Val Ile Leu Ile Arg
        780                 785                 790

CCA GAT TCA AAT TTT ATA CAG AAC CAC AAA GGT AAA TTG CTT TAT TGAAA   2514
Pro Asp Ser Asn Phe Ile Gln Asn His Lys Gly Lys Leu Leu Tyr
    795                 800                 805

CTTTCATGGT ATGAAAATGT TAGATGATAT TCAACTTATC TTATTCTTCA TCTTAATAAA   2574
AGCAGTACTT CATCATATAA AAAATAAAGT AATATACAGA TTTATACTTA CAAACTGGAC   2634
AGCAAACATG AATATGTTGA GAACTGGGAT TCTCAATTGA GGAATGGCTA CCAATATTTT   2694
GATCTGTGGT TTTGTGTTTA AGCCATGTAC TTAATTAATG ATTAACATGA GGTTACCCTA   2754
CTGTCTTTGA ACAGCGCCAC CTCTAGGCAT GCTGTCCTTG AGTTATAAGA AAGGGTACTG   2814
CATACACAAT GGACATGAAG CCAGTAATCA ACATTATTCC ACTTGCTTTC ATGGAGTTCT   2874
TACTTCCAAG TTCATGCCTT GACTTTATTC AATGTTCTAT GACAAAGGTA GAATAAATAA   2934
ATAAACACTT TCCTCACAAA AAAAAAA                                       2961
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 808 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Lys Gln Leu Cys Thr Phe Thr Ile Ser Leu Leu Phe Leu Lys Phe
 1               5                  10                  15
Ser Leu Ile Leu Cys Cys Trp Ser Glu Pro Ser Cys Phe Trp Arg Ile
            20                  25                  30
Lys Lys Ser Glu Asp Asn Asp Gly Asp Leu Gln Arg Glu Cys His Phe
        35                  40                  45
Tyr Leu Trp Lys Thr Asp Glu Pro Ile Glu Asp Ser Phe Tyr Asn Tyr
    50                  55                  60
Asp Leu Ser Phe Arg Ile Ala Gly Ser Glu Tyr Glu Leu Leu Leu Val
65                  70                  75                  80
Met Phe Phe Ala Thr Asp Glu Ile Asn Lys Asn Pro Tyr Leu Leu Pro
                85                  90                  95
Asn Met Ser Leu Met Phe Ser Ile Ile Gly Gly Asn Cys His Asp Leu
            100                 105                 110
Leu Arg Ser Leu Asp Gln Glu Tyr Ala Gln Ile Asp Gly His Met Asn
        115                 120                 125
Phe Val Asn Tyr Phe Cys Tyr Leu Asp Asp Ser Cys Ala Thr Gly Leu
    130                 135                 140
Thr Gly Pro Ser Trp Lys Thr Ser Leu Lys Leu Ala Met His Ser Ser
145                 150                 155                 160
Met Pro Leu Val Phe Phe Gly Pro Phe Asn Pro Asn Leu Arg Asp His
                165                 170                 175
Asp Arg Leu Pro His Val His Gln Val Ala Pro Lys Asp Thr His Leu
            180                 185                 190
Ser His Gly Met Val Ser Leu Met Phe His Phe Arg Trp Thr Trp Ile
        195                 200                 205
Gly Leu Val Ile Ser Asp Asp Asp Gln Gly Ile Gln Phe Leu Ser Asp
    210                 215                 220
Leu Arg Glu Glu Ser Gln Arg His Gly Ile Cys Leu Ala Phe Val Asn
225                 230                 235                 240
Met Ile Pro Glu Asn Met Gln Ile Tyr Met Thr Arg Ala Thr Ile Tyr
                245                 250                 255
Asp Thr Gln Ile Met Thr Ser Ser Ala Lys Val Val Ile Ile Tyr Gly
            260                 265                 270
Asp Met Asn Ser Thr Leu Glu Ala Ser Phe Arg Arg Trp Glu Glu Leu
        275                 280                 285
Gly Ala Arg Arg Ile Trp Ile Thr Thr Thr Gln Trp Asp Val Ile Thr
    290                 295                 300
Asn Lys Lys Asp Phe Thr Leu Asn Leu Phe His Gly Thr Ile Thr Phe
305                 310                 315                 320
Ala His His Lys Asp Glu Ile Pro Lys Phe Arg Asn Phe Met Gln Thr
                325                 330                 335
Lys Lys Thr Ala Lys Tyr Leu Val Asp Ile Ser His Thr Ile Leu Glu
            340                 345                 350
Trp Asn Tyr Phe Asn Cys Ser Ile Ser Lys Asn Ser Ser Lys Met Gly
        355                 360                 365
His Phe Thr Phe Asn Asn Thr Leu Gln Trp Thr Ala Leu His Asn Tyr
    370                 375                 380
Asp Met Ala Leu Ser Asp Glu Gly Tyr Asn Leu Tyr Asn Ala Val Tyr
385                 390                 395                 400
Ala Val Ala His Thr Tyr His Glu Tyr Ile Leu Gln Gln Val Glu Ser
                405                 410                 415
Gln Lys Lys Ala Lys Pro Lys Arg Tyr Phe Thr Ala Cys Gln Gln Val
            420                 425                 430
Ser Ser Leu Met Lys Thr Arg Val Phe Met Asn Pro Val Gly Glu Leu
        435                 440                 445
Val Asn Met Lys His Arg Glu Asn Gln Cys Thr Glu Tyr Asp Ile Phe
    450                 455                 460
```

```
Ile Ile Trp Asn Phe Pro Gln Gly Leu Gly Leu Lys Val Lys Val Gly
465                 470                 475                 480
Ser Tyr Leu Pro Cys Phe Pro Lys Ser Gln Gln Leu His Ile Ala Asp
                485                 490                 495
Asp Leu Glu Trp Ala Met Gly Gly Thr Ser Val Asp Met Glu Gln Cys
            500                 505                 510
Val Arg Cys Pro Asp Asn Lys Tyr Ala Asn Leu Glu Gln Thr His Cys
        515                 520                 525
Leu Gln Arg Thr Val Ser Phe Leu Ala Tyr Glu Asp Pro Leu Gly Met
    530                 535                 540
Ala Leu Gly Cys Met Ala Leu Ser Phe Ser Ala Ile Thr Ile Leu Val
545                 550                 555                 560
Leu Val Thr Phe Val Lys Tyr Lys Asp Thr Pro Ile Val Lys Ala Asn
                565                 570                 575
Asn Arg Ile Leu Ser Tyr Ile Leu Leu Ile Ser Leu Val Phe Cys Phe
            580                 585                 590
Leu Cys Ser Leu Leu Phe Ile Gly His Pro Asp Gln Val Thr Cys Ile
        595                 600                 605
Leu Gln Gln Thr Thr Phe Gly Val Leu Phe Thr Val Ser Val Ser Thr
    610                 615                 620
Val Leu Ala Lys Thr Ile Thr Val Val Met Ala Phe Lys Leu Thr Thr
625                 630                 635                 640
Pro Gly Arg Arg Met Arg Gly Met Met Met Thr Gly Ala Pro Lys Leu
                645                 650                 655
Val Ile Pro Ile Cys Thr Leu Ile Gln Leu Val Leu Cys Gly Ile Trp
            660                 665                 670
Leu Val Thr Ser Pro Pro Phe Ile Asp Arg Asp Ile Gln Ser Glu His
        675                 680                 685
Gly Lys Ile Val Ile Leu Cys Asn Lys Gly Ser Val Val Ala Phe His
    690                 695                 700
Val Val Leu Gly Tyr Leu Gly Ser Leu Ala Leu Gly Ser Phe Thr Leu
705                 710                 715                 720
Ala Phe Leu Ala Arg Asn Leu Pro Asp Thr Phe Asn Glu Ala Lys Phe
                725                 730                 735
Leu Thr Phe Ser Met Leu Val Phe Cys Ser Val Trp Ile Thr Phe Leu
            740                 745                 750
Pro Val Tyr His Ser Thr Arg Gly Lys Val Met Val Val Val Glu Val
        755                 760                 765
Phe Ser Ile Leu Ala Ser Ser Ala Gly Leu Leu Met Cys Ile Phe Val
    770                 775                 780
Pro Lys Cys Tyr Val Ile Leu Ile Arg Pro Asp Ser Asn Phe Ile Gln
785                 790                 795                 800
Asn His Lys Gly Lys Leu Leu Tyr
                805
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2907 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...2409
(D) OTHER INFORMATION: VR3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
CAT TTT TAC CTT GGG GCA GTT GAT AAA CCA ATT GAA GAT AAT TTT TAT      48
His Phe Tyr Leu Gly Ala Val Asp Lys Pro Ile Glu Asp Asn Phe Tyr
```

```
1               5                  10                  15

AAT TCA CTT TTA AAG TTT AGA ATT GCA GCA AGT GAA TAT GAG TTT CTT      96
Asn Ser Leu Leu Lys Phe Arg Ile Ala Ala Ser Glu Tyr Glu Phe Leu
            20                  25                  30

CTG GTA ATG TTT TTT GCT ACT GAT GAG ATC AAC AAG AAT CCT TAT CTT     144
Leu Val Met Phe Phe Ala Thr Asp Glu Ile Asn Lys Asn Pro Tyr Leu
        35                  40                  45

TTA CCC AAC ATA ACT TTG ATG TTC TCC ATC ATT GGT GGA AAC TGT CAT     192
Leu Pro Asn Ile Thr Leu Met Phe Ser Ile Ile Gly Gly Asn Cys His
    50                  55                  60

GAT TTA TTG AGA GGT TTG GAT CAA GCA TAT ACA CAA ATA AAT GGA CAT     240
Asp Leu Leu Arg Gly Leu Asp Gln Ala Tyr Thr Gln Ile Asn Gly His
65                  70                  75                  80

ATG AAT TTT GTT AAT TAT TTC TGT TAT TTA GAT GAT TCA TGT GCC ATA     288
Met Asn Phe Val Asn Tyr Phe Cys Tyr Leu Asp Asp Ser Cys Ala Ile
                85                  90                  95

GGT CTT ACA GGA CCA TCA TGG AAA ACA TCC TTA AAT CTG GCA ATG CAT     336
Gly Leu Thr Gly Pro Ser Trp Lys Thr Ser Leu Asn Leu Ala Met His
            100                 105                 110

TCT TCA ATG CCA CTG GTT TTC TTT GGA TCA TTT AAT CCT AAC CTA CAT     384
Ser Ser Met Pro Leu Val Phe Phe Gly Ser Phe Asn Pro Asn Leu His
        115                 120                 125

GAC CAT GAC CGG CTG CAC CAT GTC CAT CAA GTA GCC ACC AAG GAC ACA     432
Asp His Asp Arg Leu His His Val His Gln Val Ala Thr Lys Asp Thr
    130                 135                 140

CAT TTG TCC CAT GGC ATT GTC TCC TTG ATG TTT CAT TTT AGA TGG ACT     480
His Leu Ser His Gly Ile Val Ser Leu Met Phe His Phe Arg Trp Thr
145                 150                 155                 160

TGG ATA GGA CTG GTC ATC TCA GAT GAT GAC AAG GGT ATT CAG TTT CTC     528
Trp Ile Gly Leu Val Ile Ser Asp Asp Asp Lys Gly Ile Gln Phe Leu
                165                 170                 175

TCA GAT TTA AGA GAA GAA AGC CAA AGG CAT GGG ATC TGT TTA GCT TTT     576
Ser Asp Leu Arg Glu Glu Ser Gln Arg His Gly Ile Cys Leu Ala Phe
            180                 185                 190

GTT AAT ATG ATC CCA GAA AAC ATG CAG ATA TAC ATG ACA AGG GCT ACA     624
Val Asn Met Ile Pro Glu Asn Met Gln Ile Tyr Met Thr Arg Ala Thr
        195                 200                 205

ATA TAT GAT AAA CAA ATT ATG ACG TCT TTA GCA AAA GTT GTT ATC ATT     672
Ile Tyr Asp Lys Gln Ile Met Thr Ser Leu Ala Lys Val Val Ile Ile
    210                 215                 220

TAT GGT GAA ATG AAC TCT ACA CTA GAA GTA AGC TTT AGA AGA TGG GAA     720
Tyr Gly Glu Met Asn Ser Thr Leu Glu Val Ser Phe Arg Arg Trp Glu
225                 230                 235                 240

AAT TTA GGT GCT CGG AGA ATC TGG ATC ACA ACC TCA CAA TGG GAT GTC     768
Asn Leu Gly Ala Arg Arg Ile Trp Ile Thr Thr Ser Gln Trp Asp Val
                245                 250                 255

ATC ACA AAT AAA AAA GAA TTC ACC CTT AAT CTC TTC CAT GGG ACT ATT     816
Ile Thr Asn Lys Lys Glu Phe Thr Leu Asn Leu Phe His Gly Thr Ile
            260                 265                 270
```

```
ACT TTT GCA CAC CGC AGA TTT GAG ATT CCT AAA TTT AAA AAA TTT ATG      864
Thr Phe Ala His Arg Arg Phe Glu Ile Pro Lys Phe Lys Lys Phe Met
        275                 280                 285

CAA ACA ATG AAC ACT GCC AAA TAC CCA GTA GAT ATT TCT CAT ACT ATA      912
Gln Thr Met Asn Thr Ala Lys Tyr Pro Val Asp Ile Ser His Thr Ile
    290                 295                 300

TTG GAG TGG AAT TAT TTT AAT TGT TCA ATC TCT AAG AAC AGC AGT AAA      960
Leu Glu Trp Asn Tyr Phe Asn Cys Ser Ile Ser Lys Asn Ser Ser Lys
305                 310                 315                 320

ATG GAT CAT ATT ACA TTC AAC AAC ACA TTG GAA TGG ACA GCA CTG CAC     1008
Met Asp His Ile Thr Phe Asn Asn Thr Leu Glu Trp Thr Ala Leu His
                325                 330                 335

AAC TAT GAT ATG GTG ATG AGT GAT GAA GGT TAC AAT TTG TAT AAT GCT     1056
Asn Tyr Asp Met Val Met Ser Asp Glu Gly Tyr Asn Leu Tyr Asn Ala
            340                 345                 350

GTT TAT GCT GTG GCC CAC ACC TAC CAT GAA CAT ATT TTT CAA CAA GTA     1104
Val Tyr Ala Val Ala His Thr Tyr His Glu His Ile Phe Gln Gln Val
        355                 360                 365

GAG TCT CAG AAA AAG GCA AAA CCC AAA AGA TTT TTC ACT GTT TGT CAG     1152
Glu Ser Gln Lys Lys Ala Lys Pro Lys Arg Phe Phe Thr Val Cys Gln
    370                 375                 380

CAG GTG TCT TCC TTG ATG AAA ACC AGG GTA TTT ACT AAC CCT GTT GGA     1200
Gln Val Ser Ser Leu Met Lys Thr Arg Val Phe Thr Asn Pro Val Gly
385                 390                 395                 400

GAA CTG GTG AAC ATG AAG CAT AGG GAA AAT CAG TGT ACA GAG TAT GAC     1248
Glu Leu Val Asn Met Lys His Arg Glu Asn Gln Cys Thr Glu Tyr Asp
                405                 410                 415

ATT TTC CTC ATT TGG AAC TTT CCA CAA GGC CTT GGA TTA AAA GTG AAA     1296
Ile Phe Leu Ile Trp Asn Phe Pro Gln Gly Leu Gly Leu Lys Val Lys
            420                 425                 430

ATA GGA AGC TAT TTA CCT TGT TTT CCA CAG AGA CAA GAA CTT CAT ATA     1344
Ile Gly Ser Tyr Leu Pro Cys Phe Pro Gln Arg Gln Glu Leu His Ile
        435                 440                 445

TCT GAT GAT TTG GAA TGG GCC ATG GGA GGA ACA TCA GTG GTT CCC TCC     1392
Ser Asp Asp Leu Glu Trp Ala Met Gly Gly Thr Ser Val Val Pro Ser
    450                 455                 460

TCT GTG TGT AGT GTG GCA TGT ACT GCA GGA TTC AGG AAA ATT CAT CAG     1440
Ser Val Cys Ser Val Ala Cys Thr Ala Gly Phe Arg Lys Ile His Gln
465                 470                 475                 480

AAA GAA ACA GCA GAC TGC TGC TTT GAT TGT GTT CAG TGC CCA GAA AAT     1488
Lys Glu Thr Ala Asp Cys Cys Phe Asp Cys Val Gln Cys Pro Glu Asn
                485                 490                 495

GAG GTT TCC AAT GAA ACA GAT ATG GAA CAG TGT GTG AAG TGT CCA TAT     1536
Glu Val Ser Asn Glu Thr Asp Met Glu Gln Cys Val Lys Cys Pro Tyr
            500                 505                 510

GAT AAG TAT GCC AAC ATA GAG AAA ACC CAC TGC CTC TCA AGA GCT GTA     1584
Asp Lys Tyr Ala Asn Ile Glu Lys Thr His Cys Leu Ser Arg Ala Val
        515                 520                 525

TCA TTT CTG GCT TAT GAA GAT CCA TTG GGG ATA GCT CTA GGC TGC ATA     1632
```

```
Ser Phe Leu Ala Tyr Glu Asp Pro Leu Gly Ile Ala Leu Gly Cys Ile
    530                 535                 540

GCA CTG TCC TTC TCA GCC ATC ACA ATT CTA GTA CTA ATC ACA TTT TTG     1680
Ala Leu Ser Phe Ser Ala Ile Thr Ile Leu Val Leu Ile Thr Phe Leu
545                 550                 555                 560

AAG TAC AAG GAT ACT CCC ATT GTG AAG GCC AAT AAC CGC ATT CTC AGC     1728
Lys Tyr Lys Asp Thr Pro Ile Val Lys Ala Asn Asn Arg Ile Leu Ser
                565                 570                 575

TAC ATC CTG CTC ATC TCT CTA GTC TTC TGC TTT CTC TGC TCC CTG CTC     1776
Tyr Ile Leu Leu Ile Ser Leu Val Phe Cys Phe Leu Cys Ser Leu Leu
            580                 585                 590

TTC ATT GGA CAT CCA AAC CAG GTC TCC TGC GTC TTG CAG CAG ACC ACA     1824
Phe Ile Gly His Pro Asn Gln Val Ser Cys Val Leu Gln Gln Thr Thr
        595                 600                 605

TTT GGA GTA TTT TTC ACT GTG TCT GTT TCT ACA GTG TTG GCC AAA ACA     1872
Phe Gly Val Phe Phe Thr Val Ser Val Ser Thr Val Leu Ala Lys Thr
    610                 615                 620

ATA ACT GTG GTC ATG GCT TTC AAG CTC ACT ACT CCA GGA AGA AGA ATG     1920
Ile Thr Val Val Met Ala Phe Lys Leu Thr Thr Pro Gly Arg Arg Met
625                 630                 635                 640

AGA GAG ATG TTG GTA ACA GGG GCA CCT AAG TTG GTC ATT CCC ATT TGT     1968
Arg Glu Met Leu Val Thr Gly Ala Pro Lys Leu Val Ile Pro Ile Cys
                645                 650                 655

ACC CTA ATC CAA TTT GTT CTC TGT GGA ATC TGG TTG ATA ACA TCT CCT     2016
Thr Leu Ile Gln Phe Val Leu Cys Gly Ile Trp Leu Ile Thr Ser Pro
            660                 665                 670

CCA TTT ATT GAC AGA GAT ATA CAA TCT GAG CAT GGG AAG ATT GTC ATT     2064
Pro Phe Ile Asp Arg Asp Ile Gln Ser Glu His Gly Lys Ile Val Ile
        675                 680                 685

CTT TGC AAT AAA GGC TCT GTC ATT GCC TTC CAT GTT GTC CTG GGA TAC     2112
Leu Cys Asn Lys Gly Ser Val Ile Ala Phe His Val Val Leu Gly Tyr
    690                 695                 700

TTG GGC TCC TTG GCT CTG GGG AGC TTC ACT TTG GCT TTC TTG GCT AGG     2160
Leu Gly Ser Leu Ala Leu Gly Ser Phe Thr Leu Ala Phe Leu Ala Arg
705                 710                 715                 720

AAC CTT CCT GAC ACA TTC AAT GAA GCC AAA TTC CTG ACT TTC AGC ATG     2208
Asn Leu Pro Asp Thr Phe Asn Glu Ala Lys Phe Leu Thr Phe Ser Met
                725                 730                 735

CTG GTG TTC TGC AGT GTC TGG ATC ACC TTT CTC CCT GTC TAC CAT AGC     2256
Leu Val Phe Cys Ser Val Trp Ile Thr Phe Leu Pro Val Tyr His Ser
            740                 745                 750

ACC AGG GGG AAG GTC ATG GTG GTT GTG GAG GTT TTC TCA ATC TTG GCT     2304
Thr Arg Gly Lys Val Met Val Val Val Glu Val Phe Ser Ile Leu Ala
        755                 760                 765

TCT AGT GCA GGG TTG CTA ATG TGT ATC TTT GTC CCA AAG TGT TAT GTT     2352
Ser Ser Ala Gly Leu Leu Met Cys Ile Phe Val Pro Lys Cys Tyr Val
    770                 775                 780

ATT TTA GTT AGA CCA GAT TCA AAT TTT ATA CGG AAG TAC AAA GAT AAA     2400
Ile Leu Val Arg Pro Asp Ser Asn Phe Ile Arg Lys Tyr Lys Asp Lys
```

785 790 795 800

TTT CGT TAT TGAAATATTC ATACTATGAA AATGTTAGAT TATACTCAAC ATATTTTTC 2458
Phe Arg Tyr

TTTGTCTTAA CAAAAGTAGT ACTTAATCTT ATAAAAATTT AAATAATATA CAAATTTGAA 2518
CTTACAAACA GGACAGAACT GTCTATTGTA ATACCAATTA CAAAACTTTG GTGAAAAATG 2578
GTCTCATTCA TAAGGACACA ATTCTGAAGA TATTGAGAAC CAGGAATCTC AACTGCGGAA 2638
ACGCTACCAT CATCCTGACC TGTGGTTTTG TGTGTAAAGC ATGAACTTAA TTAATGATTA 2698
ATATAAGGTG ACCATACTGA CTGTGAACAC TACCATCTCT GGGCAAGTTG TTCTTGTAGT 2758
TGTAAGAAAA AGCTCTGAAG ACAACATGGA AGTAAAGCCA GTAATCACCA TTATCCCTCA 2818
TGCTTTCATG GAGTGGCTGC ATCCAATTTC ATGCCTTGGC TTCATTCAAT ATACTGTGAC 2878
CAAGGTACAT AAGTAAAGAA ACACTTTTC 2907

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 803 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
His Phe Tyr Leu Gly Ala Val Asp Lys Pro Ile Glu Asp Asn Phe Tyr
 1               5                  10                  15
Asn Ser Leu Leu Lys Phe Arg Ile Ala Ala Ser Glu Tyr Glu Phe Leu
            20                  25                  30
Leu Val Met Phe Phe Ala Thr Asp Glu Ile Asn Lys Asn Pro Tyr Leu
        35                  40                  45
Leu Pro Asn Ile Thr Leu Met Phe Ser Ile Ile Gly Gly Asn Cys His
    50                  55                  60
Asp Leu Leu Arg Gly Leu Asp Gln Ala Tyr Thr Gln Ile Asn Gly His
65                  70                  75                  80
Met Asn Phe Val Asn Tyr Phe Cys Tyr Leu Asp Asp Ser Cys Ala Ile
                85                  90                  95
Gly Leu Thr Gly Pro Ser Trp Lys Thr Ser Leu Asn Leu Ala Met His
            100                 105                 110
Ser Ser Met Pro Leu Val Phe Phe Gly Ser Phe Asn Pro Asn Leu His
        115                 120                 125
Asp His Asp Arg Leu His His Val His Gln Val Ala Thr Lys Asp Thr
    130                 135                 140
His Leu Ser His Gly Ile Val Ser Leu Met Phe His Phe Arg Trp Thr
145                 150                 155                 160
Trp Ile Gly Leu Val Ile Ser Asp Asp Asp Lys Gly Ile Gln Phe Leu
                165                 170                 175
Ser Asp Leu Arg Glu Glu Ser Gln Arg His Gly Ile Cys Leu Ala Phe
            180                 185                 190
Val Asn Met Ile Pro Glu Asn Met Gln Ile Tyr Met Thr Arg Ala Thr
        195                 200                 205
Ile Tyr Asp Lys Gln Ile Met Thr Ser Leu Ala Lys Val Val Ile Ile
    210                 215                 220
Tyr Gly Glu Met Asn Ser Thr Leu Glu Val Ser Phe Arg Arg Trp Glu
225                 230                 235                 240
Asn Leu Gly Ala Arg Arg Ile Trp Ile Thr Thr Ser Gln Trp Asp Val
                245                 250                 255
Ile Thr Asn Lys Lys Glu Phe Thr Leu Asn Leu Phe His Gly Thr Ile
            260                 265                 270
Thr Phe Ala His Arg Arg Phe Glu Ile Pro Lys Phe Lys Lys Phe Met
        275                 280                 285
Gln Thr Met Asn Thr Ala Lys Tyr Pro Val Asp Ile Ser His Thr Ile
```

```
    290                 295                 300
Leu Glu Trp Asn Tyr Phe Asn Cys Ser Ile Ser Lys Asn Ser Ser Lys
305                 310                 315                 320
Met Asp His Ile Thr Phe Asn Asn Thr Leu Glu Trp Thr Ala Leu His
                325                 330                 335
Asn Tyr Asp Met Val Met Ser Asp Glu Gly Tyr Asn Leu Tyr Asn Ala
            340                 345                 350
Val Tyr Ala Val Ala His Thr Tyr His Glu His Ile Phe Gln Gln Val
        355                 360                 365
Glu Ser Gln Lys Lys Ala Lys Pro Lys Arg Phe Phe Thr Val Cys Gln
    370                 375                 380
Gln Val Ser Ser Leu Met Lys Thr Arg Val Phe Thr Asn Pro Val Gly
385                 390                 395                 400
Glu Leu Val Asn Met Lys His Arg Glu Asn Gln Cys Thr Glu Tyr Asp
                405                 410                 415
Ile Phe Leu Ile Trp Asn Phe Pro Gln Gly Leu Gly Leu Lys Val Lys
            420                 425                 430
Ile Gly Ser Tyr Leu Pro Cys Phe Pro Gln Arg Gln Glu Leu His Ile
        435                 440                 445
Ser Asp Asp Leu Glu Trp Ala Met Gly Gly Thr Ser Val Val Pro Ser
    450                 455                 460
Ser Val Cys Ser Val Ala Cys Thr Ala Gly Phe Arg Lys Ile His Gln
465                 470                 475                 480
Lys Glu Thr Ala Asp Cys Cys Phe Asp Cys Val Gln Cys Pro Glu Asn
                485                 490                 495
Glu Val Ser Asn Glu Thr Asp Met Glu Gln Cys Val Lys Cys Pro Tyr
            500                 505                 510
Asp Lys Tyr Ala Asn Ile Glu Lys Thr His Cys Leu Ser Arg Ala Val
        515                 520                 525
Ser Phe Leu Ala Tyr Glu Asp Pro Leu Gly Ile Ala Leu Gly Cys Ile
    530                 535                 540
Ala Leu Ser Phe Ser Ala Ile Thr Ile Leu Val Leu Ile Thr Phe Leu
545                 550                 555                 560
Lys Tyr Lys Asp Thr Pro Ile Val Lys Ala Asn Asn Arg Ile Leu Ser
                565                 570                 575
Tyr Ile Leu Leu Ile Ser Leu Val Phe Cys Phe Leu Cys Ser Leu Leu
            580                 585                 590
Phe Ile Gly His Pro Asn Gln Val Ser Cys Val Leu Gln Gln Thr Thr
        595                 600                 605
Phe Gly Val Phe Phe Thr Val Ser Val Ser Thr Val Leu Ala Lys Thr
    610                 615                 620
Ile Thr Val Val Met Ala Phe Lys Leu Thr Thr Pro Gly Arg Arg Met
625                 630                 635                 640
Arg Glu Met Leu Val Thr Gly Ala Pro Lys Leu Val Ile Pro Ile Cys
                645                 650                 655
Thr Leu Ile Gln Phe Val Leu Cys Gly Ile Trp Leu Ile Thr Ser Pro
            660                 665                 670
Pro Phe Ile Asp Arg Asp Ile Gln Ser Glu His Gly Lys Ile Val Ile
        675                 680                 685
Leu Cys Asn Lys Gly Ser Val Ile Ala Phe His Val Val Leu Gly Tyr
    690                 695                 700
Leu Gly Ser Leu Ala Leu Gly Ser Phe Thr Leu Ala Phe Leu Ala Arg
705                 710                 715                 720
Asn Leu Pro Asp Thr Phe Asn Glu Ala Lys Phe Leu Thr Phe Ser Met
                725                 730                 735
Leu Val Phe Cys Ser Val Trp Ile Thr Phe Leu Pro Val Tyr His Ser
            740                 745                 750
Thr Arg Gly Lys Val Met Val Val Val Glu Val Phe Ser Ile Leu Ala
        755                 760                 765
Ser Ser Ala Gly Leu Leu Met Cys Ile Phe Val Pro Lys Cys Tyr Val
    770                 775                 780
Ile Leu Val Arg Pro Asp Ser Asn Phe Ile Arg Lys Tyr Lys Asp Lys
785                 790                 795                 800
Phe Arg Tyr
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3625 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 117...2672
(D) OTHER INFORMATION: VR4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
TGAATATGCA ATAAACCTCA CATTTGCACA AAGAAATAAA AGCTGGTAGA AATCTGATGT     60
GCTGATATGC ATGGCACTTC ACAATCCGCA CTGCCCAGGT TTAAGGCAGG AAAAAG ATG    119
                                                                 Met
                                                                  1

TTC ATT TTC ATG GGA GTC TTC TTC CTA CTT AAT ATT ACA CTT CTC ATG      167
Phe Ile Phe Met Gly Val Phe Phe Leu Leu Asn Ile Thr Leu Leu Met
              5                  10                  15

GCC AAT TTC ATT GAT CCC AGG TGC TTT TGG AGA ATA AAT TTG GAT GAA      215
Ala Asn Phe Ile Asp Pro Arg Cys Phe Trp Arg Ile Asn Leu Asp Glu
        20                  25                  30

ATA ACG GAT GAA TAT TTG GGA TTA TCT TGT GCT TTC ATC CTG GCA GCT      263
Ile Thr Asp Glu Tyr Leu Gly Leu Ser Cys Ala Phe Ile Leu Ala Ala
    35                  40                  45

GTT CAG ACA CCC ATT GAA AAA GAT TAT TTC AAC ACG ACT CTT AAT TTT      311
Val Gln Thr Pro Ile Glu Lys Asp Tyr Phe Asn Thr Thr Leu Asn Phe
50                  55                  60                  65

CTA AAA ACT ACT AAA AAC CAC AAA TAT GCT TTG GCA TTG GTG TTT GCA      359
Leu Lys Thr Thr Lys Asn His Lys Tyr Ala Leu Ala Leu Val Phe Ala
                70                  75                  80

ATG GAT GAA ATC AAC AGA TAT CCT GAT CTT TTA CCA AAT ATG TCT TTG      407
Met Asp Glu Ile Asn Arg Tyr Pro Asp Leu Leu Pro Asn Met Ser Leu
            85                  90                  95

ATT ATC AGA TAC TCT TTG GGC CAT TGT GAT GGA AAA ACT GTA ACA CCT      455
Ile Ile Arg Tyr Ser Leu Gly His Cys Asp Gly Lys Thr Val Thr Pro
        100                 105                 110

ACA CCA TAT TTA TTT CAT AGA AAA AAG CAA AGC CCT ATT CCT AAT TAT      503
Thr Pro Tyr Leu Phe His Arg Lys Lys Gln Ser Pro Ile Pro Asn Tyr
    115                 120                 125

TTC TGT AAT GAA GAG AGT ATG TGT TCA TTT CTG CTT TCA GGA CCC AAT      551
Phe Cys Asn Glu Glu Ser Met Cys Ser Phe Leu Leu Ser Gly Pro Asn
130                 135                 140                 145

TGG GAT GAA TCT TTA AGT TTC TGG AAG TAC CTG GAC AGC TTC TTA TCT      599
Trp Asp Glu Ser Leu Ser Phe Trp Lys Tyr Leu Asp Ser Phe Leu Ser
                150                 155                 160

CCA CGT ATC CTT CAG CTT TCC TAT GGA TCT TTC AGT TCC ATC TTC AGT      647
Pro Arg Ile Leu Gln Leu Ser Tyr Gly Ser Phe Ser Ser Ile Phe Ser
```

```
                165                 170                 175

GAT GAT GAA CAA TAT CCC TAT CTC TAT CAG ATG GCC CCA AAA GAC ACA      695
Asp Asp Glu Gln Tyr Pro Tyr Leu Tyr Gln Met Ala Pro Lys Asp Thr
        180                 185                 190

TCT CTA GCA TTG GCA ATG GTC TCC TTC ATA CTT TAT TTG AAA TGG AAT      743
Ser Leu Ala Leu Ala Met Val Ser Phe Ile Leu Tyr Leu Lys Trp Asn
    195                 200                 205

TGG ATT GGC CTT GTC ATC CCA GAT GAT GAT CAA GGA AAC CAA TTT CTT      791
Trp Ile Gly Leu Val Ile Pro Asp Asp Asp Gln Gly Asn Gln Phe Leu
210                 215                 220                 225

TTA GAG TTG AAG AAA CAG AGT GAA AAC AAA GAA ATT TGC TTT GCC TTT      839
Leu Glu Leu Lys Lys Gln Ser Glu Asn Lys Glu Ile Cys Phe Ala Phe
                230                 235                 240

GTG AAA ATG ATC TCT GTT GAT GAA GTT TCA TTT CCA CAA AAA ACT GAA      887
Val Lys Met Ile Ser Val Asp Glu Val Ser Phe Pro Gln Lys Thr Glu
            245                 250                 255

ATA AAC TAC AAA CAA ATT GTG AAG TCA CTA ACA AAT GTT ATT ATC ATT      935
Ile Asn Tyr Lys Gln Ile Val Lys Ser Leu Thr Asn Val Ile Ile Ile
        260                 265                 270

TAT GGA GAA ACA TAT AAT TTC ATT GAT TTG ATC TTC AGA ATG TGG GAA      983
Tyr Gly Glu Thr Tyr Asn Phe Ile Asp Leu Ile Phe Arg Met Trp Glu
    275                 280                 285

CCT CCC ATT TTA CAG AGA ATA TGG ATC ACC ACA AAA CAA TTG AAT TTC     1031
Pro Pro Ile Leu Gln Arg Ile Trp Ile Thr Thr Lys Gln Leu Asn Phe
290                 295                 300                 305

CCT ACC AGT AAG ACA GAC ATA AGT CAT GAC ACA TTC TAT GGA TCA CTT     1079
Pro Thr Ser Lys Thr Asp Ile Ser His Asp Thr Phe Tyr Gly Ser Leu
                310                 315                 320

ACT TTT CTA CCC CAC CAT GGT GAG ATT TCT GGC TTT AAA AAT TTT GTA     1127
Thr Phe Leu Pro His His Gly Glu Ile Ser Gly Phe Lys Asn Phe Val
            325                 330                 335

CAG ACA TGG TTC CAT CTC AGA AAC ACA GAT TTA TGT CTA GTA ATG CCA     1175
Gln Thr Trp Phe His Leu Arg Asn Thr Asp Leu Cys Leu Val Met Pro
        340                 345                 350

GAG TGG AAA TAT ATT AAC TCT GAA GAC TCA GCA TCT AAT TGT AAA ATA     1223
Glu Trp Lys Tyr Ile Asn Ser Glu Asp Ser Ala Ser Asn Cys Lys Ile
    355                 360                 365

CTT AAG AAC AGT TCA TCT GAT GCC TCA TTT GAT TGG CTA ATG GAA GAG     1271
Leu Lys Asn Ser Ser Ser Asp Ala Ser Phe Asp Trp Leu Met Glu Glu
370                 375                 380                 385

AAG CTT GAC ATG GCC TTT AGT GAG AAT AGT CAT AAC ATA TAT AAT GCT     1319
Lys Leu Asp Met Ala Phe Ser Glu Asn Ser His Asn Ile Tyr Asn Ala
                390                 395                 400

GTG CAT GCC ATA GCC CAT GCC CTC CAT GAG ATG AAT CTG CAA CAG GCT     1367
Val His Ala Ile Ala His Ala Leu His Glu Met Asn Leu Gln Gln Ala
            405                 410                 415

GAT AAT CAG GCA ATA GAT AAT GGA AAA GGA GCC AGT TCT CAC TGC TTG     1415
Asp Asn Gln Ala Ile Asp Asn Gly Lys Gly Ala Ser Ser His Cys Leu
        420                 425                 430
```

```
AAG GTA AAC TCC TTT CTA AGA AGG ACC TAC TTC ACT AAT CCT CTT GGG     1463
Lys Val Asn Ser Phe Leu Arg Arg Thr Tyr Phe Thr Asn Pro Leu Gly
    435                 440                 445

GAC AAA GTG TTT ATG AAG CAA AGA GTA ATA ATG CAG GAT GAA TAT GAC     1511
Asp Lys Val Phe Met Lys Gln Arg Val Ile Met Gln Asp Glu Tyr Asp
450                 455                 460                 465

ATT GTT CAC TTT GCG AAT CTC TCA CAA CAC CTT GGG ATT AAG ATG AAG     1559
Ile Val His Phe Ala Asn Leu Ser Gln His Leu Gly Ile Lys Met Lys
                470                 475                 480

TTA GGA AAG TTC AGC CCA TAT TTA CCA CAT GGT CGA CAC TCT CAC TTA     1607
Leu Gly Lys Phe Ser Pro Tyr Leu Pro His Gly Arg His Ser His Leu
            485                 490                 495

TAC GTA GAC ATG ATT GAG TTG GCC ACA GGA AGA AGA AAG ATG CCA TCC     1655
Tyr Val Asp Met Ile Glu Leu Ala Thr Gly Arg Arg Lys Met Pro Ser
        500                 505                 510

TCT GTG TGC AGT GCA GAT TGT AGT CCT GGA TTC AGA AGA TTA TGG AAG     1703
Ser Val Cys Ser Ala Asp Cys Ser Pro Gly Phe Arg Arg Leu Trp Lys
    515                 520                 525

GAG GGA ATG GCA GCC TGC TGT TTT GTT TGC AGC CCC TGC CCT GAA AAT     1751
Glu Gly Met Ala Ala Cys Cys Phe Val Cys Ser Pro Cys Pro Glu Asn
530                 535                 540                 545

GAA ATT TCT AAT GAG ACA AAT ATG GAT CAA TGC GTG AAT TGT CCA GAA     1799
Glu Ile Ser Asn Glu Thr Asn Met Asp Gln Cys Val Asn Cys Pro Glu
                550                 555                 560

TAC CAA TAT GCC AAC ACA GAA CAG AAC AAA TGT ATT CAG AAA GGT GTC     1847
Tyr Gln Tyr Ala Asn Thr Glu Gln Asn Lys Cys Ile Gln Lys Gly Val
            565                 570                 575

ACC TTC CTA AGC TAT GAA GAC CCC TTG GGG ATG GCA CTT GCC TTA ATG     1895
Thr Phe Leu Ser Tyr Glu Asp Pro Leu Gly Met Ala Leu Ala Leu Met
        580                 585                 590

GCC TTC TGC TTC TCT GCA TTC ACA GCT GTG GTA CTT TGT GTC TTT GTG     1943
Ala Phe Cys Phe Ser Ala Phe Thr Ala Val Val Leu Cys Val Phe Val
    595                 600                 605

AAG CAC CAT GAC ACT CCT ATT GTG AAG GCC AAT AAC AGA AGC CTC AGC     1991
Lys His His Asp Thr Pro Ile Val Lys Ala Asn Asn Arg Ser Leu Ser
610                 615                 620                 625

TAT CTA TTA CTC ATG TCA CTC ATG TTC TGT TTT CTG TGC TCC TTT TTC     2039
Tyr Leu Leu Leu Met Ser Leu Met Phe Cys Phe Leu Cys Ser Phe Phe
                630                 635                 640

TTC ATT GGC CTT CCA AAC AAA GTC ATC TGT GTC TTA CAG CAA ATC ACA     2087
Phe Ile Gly Leu Pro Asn Lys Val Ile Cys Val Leu Gln Gln Ile Thr
            645                 650                 655

TTT GGA ATT GTA TTC ACT GTG GCT GTT TCC ACA GTT CTG GCC AAA ACA     2135
Phe Gly Ile Val Phe Thr Val Ala Val Ser Thr Val Leu Ala Lys Thr
        660                 665                 670

GTC ACT GTG GTT CTA GCT TTC AAA GTC ACA GTC CCA GGA AGA AGA TTG     2183
Val Thr Val Val Leu Ala Phe Lys Val Thr Val Pro Gly Arg Arg Leu
    675                 680                 685

AGA TAC TTC CTT GTA TCA GGG ACA CTA AAC TAC ATT ATT CCT ATA TGT.    2231
```

```
Arg Tyr Phe Leu Val Ser Gly Thr Leu Asn Tyr Ile Ile Pro Ile Cys
690                 695                 700                 705

TCC CTA CTC CAA TGT GTT CTG TGT GCA ATC TGG CTA GCA GTC TCT CCT    2279
Ser Leu Leu Gln Cys Val Leu Cys Ala Ile Trp Leu Ala Val Ser Pro
                710                 715                 720

CCC TTT GTT GAT ATT GAT GAA CAC TCT CAG CAT GGC CAC ATC ATC ATT    2327
Pro Phe Val Asp Ile Asp Glu His Ser Gln His Gly His Ile Ile Ile
            725                 730                 735

GTG TGC AAC AAG GGC TCA GTT ACT GCA TTC TAC TGT GTC CTT GGA TAC    2375
Val Cys Asn Lys Gly Ser Val Thr Ala Phe Tyr Cys Val Leu Gly Tyr
        740                 745                 750

TTG GCC TGC CTG GCA CTG GGA AGC TTC ACT TTG GCT TTC TTG GCC AAG    2423
Leu Ala Cys Leu Ala Leu Gly Ser Phe Thr Leu Ala Phe Leu Ala Lys
    755                 760                 765

AAT CTG CCT GAT GCA TTC AAT GAA GCC AAG TTC TTG ACC TTC AGC ATG    2471
Asn Leu Pro Asp Ala Phe Asn Glu Ala Lys Phe Leu Thr Phe Ser Met
770                 775                 780                 785

CTA GTG TTC TGC AGT GTC TGG GTC ACC TTC CTC CCT GTG TAC CAT AGC    2519
Leu Val Phe Cys Ser Val Trp Val Thr Phe Leu Pro Val Tyr His Ser
                790                 795                 800

ACA AAG GGC AAA CAC ATG GTT GCT GTG GAG ATC TTC TCT ATC TTG GCA    2567
Thr Lys Gly Lys His Met Val Ala Val Glu Ile Phe Ser Ile Leu Ala
            805                 810                 815

TCC AGT GCA GGG ATG CTT GGA TGT ATT TTT GTA CCC AAG ATT TAT ATC    2615
Ser Ser Ala Gly Met Leu Gly Cys Ile Phe Val Pro Lys Ile Tyr Ile
        820                 825                 830

ATT TTA ATG AGA CCA GAG AGA AAT TCT ACC CAA AAG ATC AGA GAA AAA    2663
Ile Leu Met Arg Pro Glu Arg Asn Ser Thr Gln Lys Ile Arg Glu Lys
    835                 840                 845

TCA TAT TTT TGAACAAATA TTTAGGAATT CTGTCAAATG TAAAGTTGGT ACATAACCA  2721
Ser Tyr Phe
850

CCAAATATTG GGTTATAGTG CATGTGTCTA GTTTTAGAAT CACTCTCACT GGTTGCTCTA   2781
GTGATAAAAG GAAGTATCAT ATCTACTGAA CTTCCGTACA GTGTCCATAA AATCTTGCAC   2841
TCATTCACTT TCTTCATTTT CTCTCAGAGA ACTAAACTCT CTAATTATTA CAATTTTATT   2901
CTTCGTTTTG AATTTCATGG AGATTGCCCT CTGGTAACTT CCAAAAAAAC GTTGATAAGG   2961
CAGTTTAATC CACCACTTTG TGTAGAAAAA ATGAGATCTA GGACAGACAG GGTTACACAT   3021
AGAAACCATC TACCAAATCA AATAATCAAT GAGAAACACA GACTAACTAA ATAATCAGCA   3081
AAGTTGAAAT CAGAACTTAT TTTCTGATTT CCAGTAAGAG CACACACAGA AGAAAATACT   3141
GACTTTTTTT TTCTTCTGTT CTTCAAGCTA CTGGCCAATA ATCTAAGGAG GAAATGTTCC   3201
TTTTCTGCTG TCAAATACAA ATATATTATA TCCAACAATG ATCAGAAGCC CAGGGATTCT   3261
GTGGCTGAAT TGGGAATATT TGGAAGAAGC TGAGGAGGAG GGTGACCAGC ATTCTCAACA   3321
AACCTGGACA AGCAAGATCT CTCAGACACT GAGCCTCTAA CCAGAGATCA TACACAAGCT   3381
GATGTGAAGC CCCCAACAAA TATGCACCAT AAGACTGCCT GGTCTAGCAT CAGTGGGAGA   3441
CACACCTAAC CCCAGAGAGA CTTAAGTCCC CAGGGATTGG GAAGTGCTGG GCATTGGGGA   3501
TGTAGGGATA TCATCTTGGA GATGGCAGAG GAGTTGTTAG ATGAGGAAGA GTCAGTGGGG   3561
CAAACCAGGA GGGGGATAAC TACTAGATTG TAACAAAAAT ATTGAGTAAT AATAAATTAA   3621
AAAA                                                               3625
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 852 amino acids
(B) TYPE: amino acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Met Phe Ile Phe Met Gly Val Phe Phe Leu Leu Asn Ile Thr Leu Leu
 1               5                  10                  15
Met Ala Asn Phe Ile Asp Pro Arg Cys Phe Trp Arg Ile Asn Leu Asp
            20                  25                  30
Glu Ile Thr Asp Glu Tyr Leu Gly Leu Ser Cys Ala Phe Ile Leu Ala
        35                  40                  45
Ala Val Gln Thr Pro Ile Glu Lys Asp Tyr Phe Asn Thr Thr Leu Asn
    50                  55                  60
Phe Leu Lys Thr Thr Lys Asn His Lys Tyr Ala Leu Ala Leu Val Phe
65                  70                  75                  80
Ala Met Asp Glu Ile Asn Arg Tyr Pro Asp Leu Leu Pro Asn Met Ser
                85                  90                  95
Leu Ile Ile Arg Tyr Ser Leu Gly His Cys Asp Gly Lys Thr Val Thr
            100                 105                 110
Pro Thr Pro Tyr Leu Phe His Arg Lys Lys Gln Ser Pro Ile Pro Asn
        115                 120                 125
Tyr Phe Cys Asn Glu Glu Ser Met Cys Ser Phe Leu Leu Ser Gly Pro
    130                 135                 140
Asn Trp Asp Glu Ser Leu Ser Phe Trp Lys Tyr Leu Asp Ser Phe Leu
145                 150                 155                 160
Ser Pro Arg Ile Leu Gln Leu Ser Tyr Gly Ser Phe Ser Ser Ile Phe
                165                 170                 175
Ser Asp Asp Glu Gln Tyr Pro Tyr Leu Tyr Gln Met Ala Pro Lys Asp
            180                 185                 190
Thr Ser Leu Ala Leu Ala Met Val Ser Phe Ile Leu Tyr Leu Lys Trp
        195                 200                 205
Asn Trp Ile Gly Leu Val Ile Pro Asp Asp Asp Gln Gly Asn Gln Phe
    210                 215                 220
Leu Leu Glu Leu Lys Lys Gln Ser Glu Asn Lys Glu Ile Cys Phe Ala
225                 230                 235                 240
Phe Val Lys Met Ile Ser Val Asp Glu Val Ser Phe Pro Gln Lys Thr
                245                 250                 255
Glu Ile Asn Tyr Lys Gln Ile Val Lys Ser Leu Thr Asn Val Ile Ile
            260                 265                 270
Ile Tyr Gly Glu Thr Tyr Asn Phe Ile Asp Leu Ile Phe Arg Met Trp
        275                 280                 285
Glu Pro Pro Ile Leu Gln Arg Ile Trp Ile Thr Thr Lys Gln Leu Asn
    290                 295                 300
Phe Pro Thr Ser Lys Thr Asp Ile Ser His Asp Thr Phe Tyr Gly Ser
305                 310                 315                 320
Leu Thr Phe Leu Pro His His Gly Glu Ile Ser Gly Phe Lys Asn Phe
                325                 330                 335
Val Gln Thr Trp Phe His Leu Arg Asn Thr Asp Leu Cys Leu Val Met
            340                 345                 350
Pro Glu Trp Lys Tyr Ile Asn Ser Glu Asp Ser Ala Ser Asn Cys Lys
        355                 360                 365
Ile Leu Lys Asn Ser Ser Ser Asp Ala Ser Phe Asp Trp Leu Met Glu
    370                 375                 380
Glu Lys Leu Asp Met Ala Phe Ser Glu Asn Ser His Asn Ile Tyr Asn
385                 390                 395                 400
Ala Val His Ala Ile Ala His Ala Leu His Glu Met Asn Leu Gln Gln
                405                 410                 415
Ala Asp Asn Gln Ala Ile Asp Asn Gly Lys Gly Ala Ser Ser His Cys
            420                 425                 430
Leu Lys Val Asn Ser Phe Leu Arg Arg Thr Tyr Phe Thr Asn Pro Leu
        435                 440                 445
Gly Asp Lys Val Phe Met Lys Gln Arg Val Ile Met Gln Asp Glu Tyr
```

```
    450                 455                 460
Asp Ile Val His Phe Ala Asn Leu Ser Gln His Leu Gly Ile Lys Met
465                 470                 475                 480
Lys Leu Gly Lys Phe Ser Pro Tyr Leu Pro His Gly Arg His Ser His
                485                 490                 495
Leu Tyr Val Asp Met Ile Glu Leu Ala Thr Gly Arg Arg Lys Met Pro
            500                 505                 510
Ser Ser Val Cys Ser Ala Asp Cys Ser Pro Gly Phe Arg Arg Leu Trp
        515                 520                 525
Lys Glu Gly Met Ala Ala Cys Cys Phe Val Cys Ser Pro Cys Pro Glu
    530                 535                 540
Asn Glu Ile Ser Asn Glu Thr Asn Met Asp Gln Cys Val Asn Cys Pro
545                 550                 555                 560
Glu Tyr Gln Tyr Ala Asn Thr Glu Gln Asn Lys Cys Ile Gln Lys Gly
                565                 570                 575
Val Thr Phe Leu Ser Tyr Glu Asp Pro Leu Gly Met Ala Leu Ala Leu
            580                 585                 590
Met Ala Phe Cys Phe Ser Ala Phe Thr Ala Val Val Leu Cys Val Phe
        595                 600                 605
Val Lys His His Asp Thr Pro Ile Val Lys Ala Asn Asn Arg Ser Leu
    610                 615                 620
Ser Tyr Leu Leu Leu Met Ser Leu Met Phe Cys Phe Leu Cys Ser Phe
625                 630                 635                 640
Phe Phe Ile Gly Leu Pro Asn Lys Val Ile Cys Val Leu Gln Gln Ile
                645                 650                 655
Thr Phe Gly Ile Val Phe Thr Val Ala Val Ser Thr Val Leu Ala Lys
            660                 665                 670
Thr Val Thr Val Val Leu Ala Phe Lys Val Thr Val Pro Gly Arg Arg
        675                 680                 685
Leu Arg Tyr Phe Leu Val Ser Gly Thr Leu Asn Tyr Ile Ile Pro Ile
    690                 695                 700
Cys Ser Leu Leu Gln Cys Val Leu Cys Ala Ile Trp Leu Ala Val Ser
705                 710                 715                 720
Pro Pro Phe Val Asp Ile Asp Glu His Ser Gln His Gly His Ile Ile
                725                 730                 735
Ile Val Cys Asn Lys Gly Ser Val Thr Ala Phe Tyr Cys Val Leu Gly
            740                 745                 750
Tyr Leu Ala Cys Leu Ala Leu Gly Ser Phe Thr Leu Ala Phe Leu Ala
        755                 760                 765
Lys Asn Leu Pro Asp Ala Phe Asn Glu Ala Lys Phe Leu Thr Phe Ser
    770                 775                 780
Met Leu Val Phe Cys Ser Val Trp Val Thr Phe Leu Pro Val Tyr His
785                 790                 795                 800
Ser Thr Lys Gly Lys His Met Val Ala Val Glu Ile Phe Ser Ile Leu
                805                 810                 815
Ala Ser Ser Ala Gly Met Leu Gly Cys Ile Phe Val Pro Lys Ile Tyr
            820                 825                 830
Ile Ile Leu Met Arg Pro Glu Arg Asn Ser Thr Gln Lys Ile Arg Glu
        835                 840                 845
Lys Ser Tyr Phe
    850
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3125 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...2169

(D) OTHER INFORMATION: VR5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
ATC TGT AAT GAA GAG AGT ATG TGT TCA TTT CTG CTT TCA GGA CCC AAT      48
Ile Cys Asn Glu Glu Ser Met Cys Ser Phe Leu Leu Ser Gly Pro Asn
 1               5                  10                  15

TGG GAT GAA TCT TTA AGT TTC TGG AAG TAC CTG GAC AGC TTC TTA TCT      96
Trp Asp Glu Ser Leu Ser Phe Trp Lys Tyr Leu Asp Ser Phe Leu Ser
            20                  25                  30

CCA CAT ATC CTT CAG CTT TCC TAT GGA TCT TTC AGT TCC ATC TTC AGT     144
Pro His Ile Leu Gln Leu Ser Tyr Gly Ser Phe Ser Ser Ile Phe Ser
        35                  40                  45

GAT GAT GAA CAA TAT CCC TAT CTC TAT CAG ATG GCC CCA AAG GAC ACA     192
Asp Asp Glu Gln Tyr Pro Tyr Leu Tyr Gln Met Ala Pro Lys Asp Thr
    50                  55                  60

TCT CTA GCA TTG GCA ATG GTC TCC TTC ATA CTT TAT TTG AAA TGG AAT     240
Ser Leu Ala Leu Ala Met Val Ser Phe Ile Leu Tyr Leu Lys Trp Asn
65                  70                  75                  80

TGG ATT GGC CTT GTC ATC CCA GAT GAC GAT CAA GGA AAC CAA TTT CTT     288
Trp Ile Gly Leu Val Ile Pro Asp Asp Asp Gln Gly Asn Gln Phe Leu
                85                  90                  95

TTA GAG TTG AAG AAA CAG AGT GAA AAC AAA GAA ATT TGC TTT GCC TTT     336
Leu Glu Leu Lys Lys Gln Ser Glu Asn Lys Glu Ile Cys Phe Ala Phe
            100                 105                 110

GTG AAA ATG ATA TCT GTT GAT GAA GTT TCA TTT CCA CAA AAA ACT GAA     384
Val Lys Met Ile Ser Val Asp Glu Val Ser Phe Pro Gln Lys Thr Glu
        115                 120                 125

ATA TAC TAC AAA CAA ATT GTG AAG TCA TTA ACA AAT GTT ATT ATC ATT     432
Ile Tyr Tyr Lys Gln Ile Val Lys Ser Leu Thr Asn Val Ile Ile Ile
    130                 135                 140

TAT GGA GAA ACA TAT AAT TTC ATT GAT TTG ATC TTC AGA ATG TGG GAA     480
Tyr Gly Glu Thr Tyr Asn Phe Ile Asp Leu Ile Phe Arg Met Trp Glu
145                 150                 155                 160

CCT CCC ATT TTA CAG AGA ATA TGG ATC ACC ACA AAA CAA TTG AAT TTC     528
Pro Pro Ile Leu Gln Arg Ile Trp Ile Thr Thr Lys Gln Leu Asn Phe
                165                 170                 175

CCT ACC AGT AAG ACA GAC ATA AGT CAT GAC ACA TTC TAT GGA TCA CTT     576
Pro Thr Ser Lys Thr Asp Ile Ser His Asp Thr Phe Tyr Gly Ser Leu
            180                 185                 190

ACT TTT CTA CCC CAC CAT GGT GAG ATT TCT GGC TTT AAA AAT TTT GTA     624
Thr Phe Leu Pro His His Gly Glu Ile Ser Gly Phe Lys Asn Phe Val
        195                 200                 205

CAG ACA TGG TTC CAT CTC AGA AAC ACA GAT TTA TAT CTA GTA ATG CCA     672
Gln Thr Trp Phe His Leu Arg Asn Thr Asp Leu Tyr Leu Val Met Pro
    210                 215                 220

GAG TGG AAA TAT ATT AAC TCT GAA GAC TCA GCA TCT AAT TGT AAA ATA     720
Glu Trp Lys Tyr Ile Asn Ser Glu Asp Ser Ala Ser Asn Cys Lys Ile
225                 230                 235                 240
```

```
CTG AAG AAC AGT TCA TCT GAT GCC TCA TTT GAT TGG CTA ATG GAA CAG      768
Leu Lys Asn Ser Ser Ser Asp Ala Ser Phe Asp Trp Leu Met Glu Gln
                245                 250                 255

AAG CTT GAC ATG GCC TTT AGT GAT AAT AGT CAT AAC ATA TAT AAT GTT      816
Lys Leu Asp Met Ala Phe Ser Asp Asn Ser His Asn Ile Tyr Asn Val
            260                 265                 270

GTG CAT GCC ATA GCC CAT GCC CTC CAT GAG ATG AAT CTG CAA CAG GCT      864
Val His Ala Ile Ala His Ala Leu His Glu Met Asn Leu Gln Gln Ala
        275                 280                 285

GAT AAT CAG GCA ATA GAT AAT GGA AAA GGA GCC AGT TCT CAC TGC TTG      912
Asp Asn Gln Ala Ile Asp Asn Gly Lys Gly Ala Ser Ser His Cys Leu
    290                 295                 300

AAG GTA AAC TCC TTT CTA AGA AGG ACC TAC TTC ACT AAT CCT CTT GGG      960
Lys Val Asn Ser Phe Leu Arg Arg Thr Tyr Phe Thr Asn Pro Leu Gly
305                 310                 315                 320

GAC AAA GTG TTT ATG AAG CAA AGA GTA ATA ATG CAG GAT GAA TAT GAC     1008
Asp Lys Val Phe Met Lys Gln Arg Val Ile Met Gln Asp Glu Tyr Asp
                325                 330                 335

ATT GTT CAC TTT GCG AAT CTC TCA CAA CAC CTT GGG ATT AAG ATG AAG     1056
Ile Val His Phe Ala Asn Leu Ser Gln His Leu Gly Ile Lys Met Lys
            340                 345                 350

TTA GGA AAG TTC AGC CCA TAT TTA CCA CAT GGT CGA CAC TCT CAC TTA     1104
Leu Gly Lys Phe Ser Pro Tyr Leu Pro His Gly Arg His Ser His Leu
        355                 360                 365

TAC GTA GAC ATG ATT GAG TTG GCC ACA GGA AGA AGA AAG ATG CCA TCC     1152
Tyr Val Asp Met Ile Glu Leu Ala Thr Gly Arg Arg Lys Met Pro Ser
    370                 375                 380

TCT GTG TGC AGT GCA GAT TGT AGT CCT GGA TTC AGA AGA TTA TGG AAG     1200
Ser Val Cys Ser Ala Asp Cys Ser Pro Gly Phe Arg Arg Leu Trp Lys
385                 390                 395                 400

GAG GGA ATG GCA GCC TGC TGT TTT GTT TGC AGC CCC TGC CCT GAA AAT     1248
Glu Gly Met Ala Ala Cys Cys Phe Val Cys Ser Pro Cys Pro Glu Asn
                405                 410                 415

GAA ATT TCT AAT GAG ACA AAT ATG GAT CAA TGC GTG AAT TGT CCA GAA     1296
Glu Ile Ser Asn Glu Thr Asn Met Asp Gln Cys Val Asn Cys Pro Glu
            420                 425                 430

TAC CAA TAT GCC AAC ACA GAA CAG AAC AAA TGT ATT CAG AAA GGT GTC     1344
Tyr Gln Tyr Ala Asn Thr Glu Gln Asn Lys Cys Ile Gln Lys Gly Val
        435                 440                 445

ACC TTC CTA AGC TAT GAA GAC CCC TTG GGG ATG GCA CTT GCC TTA ATG     1392
Thr Phe Leu Ser Tyr Glu Asp Pro Leu Gly Met Ala Leu Ala Leu Met
    450                 455                 460

GCC TTC TGC TTC TCT GCA TTC ACA GCT GTG GTA CTT TGT GTC TTT GTG     1440
Ala Phe Cys Phe Ser Ala Phe Thr Ala Val Val Leu Cys Val Phe Val
465                 470                 475                 480

AAG CAC CAT GAC ACT CCT ATT GTG AAG GCC AAT AAC AGA AGC CTC AGC     1488
Lys His His Asp Thr Pro Ile Val Lys Ala Asn Asn Arg Ser Leu Ser
                485                 490                 495

TAT CTA TTA CTC ATG TCA CTC ATG TTC TGT TTT CTG TGC TCC TTT TTC     1536
```

```
Tyr Leu Leu Leu Met Ser Leu Met Phe Cys Phe Leu Cys Ser Phe Phe
                500                 505                 510

TTC ATT GGC CTT CCA AAC AAA GTC ATC TGT GTC TTA CAG CAG ATC ACA    1584
Phe Ile Gly Leu Pro Asn Lys Val Ile Cys Val Leu Gln Gln Ile Thr
        515                 520                 525

TTT GGA ATT GTA TTT ACT GTA GCT GTT TCC ACA GTT CTG GCC AAA ACA    1632
Phe Gly Ile Val Phe Thr Val Ala Val Ser Thr Val Leu Ala Lys Thr
    530                 535                 540

GTC ACT GTG GTT CTA GCT TTC AAA GTC ACA GAC CCA GGA AGA AGA TTG    1680
Val Thr Val Val Leu Ala Phe Lys Val Thr Asp Pro Gly Arg Arg Leu
545                 550                 555                 560

AGA TAC TTC CTT GTA TCA GGG ACA CTA AAC TAC ATT ATT CCT ATA TGT    1728
Arg Tyr Phe Leu Val Ser Gly Thr Leu Asn Tyr Ile Ile Pro Ile Cys
                565                 570                 575

TCC CTA CTC CAA TGT GTT CTG TGT GCA ATC TGG CTA GCA GTC TCT CCT    1776
Ser Leu Leu Gln Cys Val Leu Cys Ala Ile Trp Leu Ala Val Ser Pro
            580                 585                 590

CCC TTT GTT GAT ATT GAT GAA CAC TCT CAG CAT GGC CAC ATC ATC ATT    1824
Pro Phe Val Asp Ile Asp Glu His Ser Gln His Gly His Ile Ile Ile
        595                 600                 605

GTG TGC AAC AAG GGC TCA GTT ACT GCA TTC TAC TGT GTC CTT GGA TAC    1872
Val Cys Asn Lys Gly Ser Val Thr Ala Phe Tyr Cys Val Leu Gly Tyr
    610                 615                 620

TTG GCC TGC CTG GCA CTG GGA AGC TTC ACT TTG GCT TTC TTG GCC AAG    1920
Leu Ala Cys Leu Ala Leu Gly Ser Phe Thr Leu Ala Phe Leu Ala Lys
625                 630                 635                 640

AAT CTG CCT GAT GCA TTC AAT GAA GCC AAG TTC TTG ACC TTC AGC ATG    1968
Asn Leu Pro Asp Ala Phe Asn Glu Ala Lys Phe Leu Thr Phe Ser Met
                645                 650                 655

CTA GTG TTC TGC AGT GTC TGG GTC ACC TTC CTC CCT GTG TAC CAT AGC    2016
Leu Val Phe Cys Ser Val Trp Val Thr Phe Leu Pro Val Tyr His Ser
            660                 665                 670

ACA AAG GGC AAA CAC ATG GTT GCT GTG GAG ATC TTC TCC ATC TTG GCA    2064
Thr Lys Gly Lys His Met Val Ala Val Glu Ile Phe Ser Ile Leu Ala
        675                 680                 685

TCC AGT GCA GGG ATG CTT GAA TGT ATT TTT GTA CCC AAG ATT TAT ATC    2112
Ser Ser Ala Gly Met Leu Glu Cys Ile Phe Val Pro Lys Ile Tyr Ile
    690                 695                 700

ATT TTA ATG AGA CCA GAG AGA AAT TCT ACC CAA AAG ATC AGG GAA AAA    2160
Ile Leu Met Arg Pro Glu Arg Asn Ser Thr Gln Lys Ile Arg Glu Lys
705                 710                 715                 720

TCA TAT TTC TGAACAAATA TTTAGGAATT CTGTCAAATG TAAAGTTGGT ACATAACCA  2218
Ser Tyr Phe

CCAAATATTG GGTTATAGTG CATGTGTCTA GTTTTAGAAT CACTCTCACT GGTTGCTCTA   2278
GTGATAAAAG GAAGTATCAT ATCTACTGAA CTTATGTACA GTGTCCATAA AATCTTGCAC   2338
TCATTCACTT TCTTCATTTT CTCTCAGAGA ACTAAACTCT CTAATTATTA CAATTTTATT   2398
CTTCGTTTTG ATTTCATGGA GATTGCCCTC TGGTAACTTC CAAAAACCGT TGATAAGGCA   2458
GTTTAATCCA CCACTTTGTG TAGAAAAAAT GAGATCTAGG ACAGACAGGG TTACACATAG   2518
AAACCATCTA CCAAATCAAA TAATCAATGA GAAACACAGA CTAACTAAAT AATCAGCAAA   2578
```

```
GTTGAAATCA GAATTATTTT CTGATTTCCA GTAAGAGCAC ACACAGAAGA AAATACTGAC   2638
TTTTTTTTTC TTCTGTTCTT CAAGCTACTG GCCAATAATC TAAGGAGGAA ATGTTCCTTT   2698
TCTGCTGTCA AATACAAATA TATTATATCC AACAATGATC AGAAGCCCAG GGATTCTGTG   2758
GCTGAATTGG GAATATTTGG AAGAAGCTGA GGAGGAGGGT GACCAGCATT CTCAACAAAC   2818
CTGGACAAGC AAGATCTCTC AGACACTGAG CCTCTAACCA GAGATCATAC ACAAGCTGAT   2878
GTGAAGCCCC CAACAAATAT GCACCATAAG ACTGCCTGGT CTAGCATCAG TGGGAGACAC   2938
ACCTAACCCC AGAGAGACTT AAGTCCCCAG GGATTGGGAA GTGCTGGGCA TTGAGGATGT   2998
AGGGATATCA TCTTTGAGAT GGCAGAGGAG TTGTTAGATG AGGAAGAGTC AGGGGGGCAA   3058
ACCAGGAAGG GGATAACTAC TAGATTGTAA CAAAAATATT GAGTAATAAT AAATTAAAAA   3118
ATGAAAT                                                             3125
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 723 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Ile Cys Asn Glu Glu Ser Met Cys Ser Phe Leu Leu Ser Gly Pro Asn
 1               5                  10                  15
Trp Asp Glu Ser Leu Ser Phe Trp Lys Tyr Leu Asp Ser Phe Leu Ser
            20                  25                  30
Pro His Ile Leu Gln Leu Ser Tyr Gly Ser Phe Ser Ser Ile Phe Ser
        35                  40                  45
Asp Asp Glu Gln Tyr Pro Tyr Leu Tyr Gln Met Ala Pro Lys Asp Thr
    50                  55                  60
Ser Leu Ala Leu Ala Met Val Ser Phe Ile Leu Tyr Leu Lys Trp Asn
65                  70                  75                  80
Trp Ile Gly Leu Val Ile Pro Asp Asp Asp Gln Gly Asn Gln Phe Leu
                85                  90                  95
Leu Glu Leu Lys Lys Gln Ser Glu Asn Lys Glu Ile Cys Phe Ala Phe
            100                 105                 110
Val Lys Met Ile Ser Val Asp Glu Val Ser Phe Pro Gln Lys Thr Glu
        115                 120                 125
Ile Tyr Tyr Lys Gln Ile Val Lys Ser Leu Thr Asn Val Ile Ile Ile
    130                 135                 140
Tyr Gly Glu Thr Tyr Asn Phe Ile Asp Leu Ile Phe Arg Met Trp Glu
145                 150                 155                 160
Pro Pro Ile Leu Gln Arg Ile Trp Ile Thr Thr Lys Gln Leu Asn Phe
                165                 170                 175
Pro Thr Ser Lys Thr Asp Ile Ser His Asp Thr Phe Tyr Gly Ser Leu
            180                 185                 190
Thr Phe Leu Pro His His Gly Glu Ile Ser Gly Phe Lys Asn Phe Val
        195                 200                 205
Gln Thr Trp Phe His Leu Arg Asn Thr Asp Leu Tyr Leu Val Met Pro
    210                 215                 220
Glu Trp Lys Tyr Ile Asn Ser Glu Asp Ser Ala Ser Asn Cys Lys Ile
225                 230                 235                 240
Leu Lys Asn Ser Ser Ser Asp Ala Ser Phe Asp Trp Leu Met Glu Gln
                245                 250                 255
Lys Leu Asp Met Ala Phe Ser Asp Asn Ser His Asn Ile Tyr Asn Val
            260                 265                 270
Val His Ala Ile Ala His Ala Leu His Glu Met Asn Leu Gln Gln Ala
        275                 280                 285
Asp Asn Gln Ala Ile Asp Asn Gly Lys Gly Ala Ser Ser His Cys Leu
    290                 295                 300
Lys Val Asn Ser Phe Leu Arg Arg Thr Tyr Phe Thr Asn Pro Leu Gly
305                 310                 315                 320
Asp Lys Val Phe Met Lys Gln Arg Val Ile Met Gln Asp Glu Tyr Asp
```

```
                 325                 330                 335
Ile Val His Phe Ala Asn Leu Ser Gln His Leu Gly Ile Lys Met Lys
            340                 345                 350
Leu Gly Lys Phe Ser Pro Tyr Leu Pro His Gly Arg His Ser His Leu
        355                 360                 365
Tyr Val Asp Met Ile Glu Leu Ala Thr Gly Arg Arg Lys Met Pro Ser
    370                 375                 380
Ser Val Cys Ser Ala Asp Cys Ser Pro Gly Phe Arg Arg Leu Trp Lys
385                 390                 395                 400
Glu Gly Met Ala Ala Cys Cys Phe Val Cys Ser Pro Cys Pro Glu Asn
                405                 410                 415
Glu Ile Ser Asn Glu Thr Asn Met Asp Gln Cys Val Asn Cys Pro Glu
            420                 425                 430
Tyr Gln Tyr Ala Asn Thr Glu Gln Asn Lys Cys Ile Gln Lys Gly Val
        435                 440                 445
Thr Phe Leu Ser Tyr Glu Asp Pro Leu Gly Met Ala Leu Ala Leu Met
    450                 455                 460
Ala Phe Cys Phe Ser Ala Phe Thr Ala Val Val Leu Cys Val Phe Val
465                 470                 475                 480
Lys His His Asp Thr Pro Ile Val Lys Ala Asn Asn Arg Ser Leu Ser
                485                 490                 495
Tyr Leu Leu Leu Met Ser Leu Met Phe Cys Phe Leu Cys Ser Phe Phe
            500                 505                 510
Phe Ile Gly Leu Pro Asn Lys Val Ile Cys Val Leu Gln Gln Ile Thr
        515                 520                 525
Phe Gly Ile Val Phe Thr Val Ala Val Ser Thr Val Leu Ala Lys Thr
    530                 535                 540
Val Thr Val Val Leu Ala Phe Lys Val Thr Asp Pro Gly Arg Arg Leu
545                 550                 555                 560
Arg Tyr Phe Leu Val Ser Gly Thr Leu Asn Tyr Ile Ile Pro Ile Cys
                565                 570                 575
Ser Leu Leu Gln Cys Val Leu Cys Ala Ile Trp Leu Ala Val Ser Pro
            580                 585                 590
Pro Phe Val Asp Ile Asp Glu His Ser Gln His Gly His Ile Ile Ile
        595                 600                 605
Val Cys Asn Lys Gly Ser Val Thr Ala Phe Tyr Cys Val Leu Gly Tyr
    610                 615                 620
Leu Ala Cys Leu Ala Leu Gly Ser Phe Thr Leu Ala Phe Leu Ala Lys
625                 630                 635                 640
Asn Leu Pro Asp Ala Phe Asn Glu Ala Lys Phe Leu Thr Phe Ser Met
                645                 650                 655
Leu Val Phe Cys Ser Val Trp Val Thr Phe Leu Pro Val Tyr His Ser
            660                 665                 670
Thr Lys Gly Lys His Met Val Ala Val Glu Ile Phe Ser Ile Leu Ala
        675                 680                 685
Ser Ser Ala Gly Met Leu Glu Cys Ile Phe Val Pro Lys Ile Tyr Ile
    690                 695                 700
Ile Leu Met Arg Pro Glu Arg Asn Ser Thr Gln Lys Ile Arg Glu Lys
705                 710                 715                 720
Ser Tyr Phe
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1889 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
GAATTCGGCT TCTGCACCAA ATGGCGACGA AAGACACATC TCTTTCACTT GCCATTGTTT     60
CTTTGATGGT TCATTTTAGG TGGTCTTGGG TTGGTCTAAT TCTCCCAGAT GACCACAAAG    120
GAAATAAAAT ACTATCAGAT TTTAGAAAGG AGATGGAAAG AAAAAGAATC TGTACGGCTT    180
TTGTAAAAAT GATTCCTGCC ACATGGACTT CATCTTTTGT CAAATTCTGG GAAAATATGG    240
ATGACACCAA CATAATAATT ATTTATGGTG ACATTGATTC TCTAGAAGGT CTAATGCGAA    300
ATATTGGGCA AAGGTTATTG ACATGGCATG TCTGGGTCAT GAACATTGAA CCCCATATTA    360
TTGAATATGA TAATTATTTC ATGTTAGATT CATTCCATGG AAGTTTAATT TTTAAGCACA    420
ATTATAGAGA GAATTTTGAG TTTACCAAAT TTATTCGAAC AGTTAATCCT AAAAAATACC    480
CAGAAGACAT TTATCTCCCT AAGATGTGGT ATTTGTTCTT CATGTGCTCA TTTTCTGATA    540
TTAATTGTCA AGTTTTGGAC AGCTGTCAAA CAAATGCTTC TTTGGATATG TTACCTAGTC    600
AGATATTTGA TGTGGTCATG AGTGAAGAGA GCACAAGTAT TTACAATGCT GTGTACGCTG    660
TGGCTCACAG CCTCCATGAG ATGAGACTTC AGCAACTTCA AACACAACCG TGTGAAAATG    720
AAGAAGGGAT GGAGTTCTTT CCATGGCAGC TTAATACTTT CCTGAAGGAT ATTGAGGTGA    780
GAGTCAACAG TTTAGACTGG AGACAGAGAA TAGATGCTGA ATATGACATT CTTAACCTCT    840
GGAATTTACC AAAGGGTCTT GGACTAAAAG TGAAAATAGG AAACTTTTAT GCAAATGCTC    900
CCCAGGGTCA ACAATTGTCT TTATCTGAAC AGATGATTCA ATGGCCAGAA ATATTTTCAG    960
AGATCCCTCA GTCGGTGTGC AGTGAGAGTT GTGGGCCTGG ATTCAGGAAA GTAACCCTGG   1020
AGAATAAGGC TATCTGCTGC TACAATTGTA CTCCCTGTGC AGACAATGAG ATTTCTAATG   1080
AGACAGATGT AGACCAGTGT GTGAAGTGTC CAGAGAGTCA TTATGCAAAT ACAGAGAAGA   1140
GCAACTGCTA TCAAAAGTCT GTGAGCTTTC TGGGCTATGA AGACCCTTTG GGGATGGCTC   1200
TAGCCAGCAT AGCTTTGTGC TTGTCTGCAC TAACTGCCTT TGTTATTGGC ATATTTGTGA   1260
AACACAAAGA CACTCCTATT GTTAAGGCCA ATAATCAAGC TCTGAGTTAC ACTTTGCTCA   1320
TCACACTCAA ATTCTGTTTC CTATGTTCTT TGAACTTCAT TGGTCAGCCC AACACAGTTG   1380
CCTGCATCCT TCAGCAGACC ACCTTTGCAG TTGCTTTCAC TATGGCTCTT GCCACTGTGT   1440
TGGCCAAAGC TATCACTGTG GTTCTTGCCT TTAAGGTCAG TTTTCCAGGG AGAATGGTAA   1500
GATGGCTAAT GATATCAAGG GGTCCAAACT ATATCATTCC TATCTGCACC CTGATCCAAC   1560
TTCTTCTTTG TGGAATATGG ATGGCAATAT CTCCACCATA CATTGACCAA GATGCTCATA   1620
TTGAACATGG TCACATCATC ATTTTGTGCA ACAAGGGCTC AGCTGTTGCC TTCCACTCTG   1680
TCCTGGGATA CCTCTGCTTC TTGGCCCTTG GGAGTTATAC CATGGCCTTC TTGTCAAGAA   1740
ATTTGCCTGA TACATTCAAC GAATCCAAAT TTATCTCACT AAGTATGCTG GTATTCTTCT   1800
GTGTCTGGAT CACCTTTCTT CCTGTCTACC ACAGCACTAA AGGGAAGGTC ATGGTCGCCG   1860
TCGAGGTCTT TTGCATCCAA GCCGAATTC                                     1889
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 604 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Ser Leu Ser Leu Ala Ile Val Ser Leu Met Val His Phe Arg Trp Ser
 1               5                  10                  15
Trp Val Gly Leu Ile Leu Pro Asp Asp His Lys Gly Asn Lys Ile Leu
            20                  25                  30
Ser Asp Phe Arg Lys Glu Met Glu Arg Lys Arg Ile Cys Thr Ala Phe
        35                  40                  45
Val Lys Met Ile Pro Ala Thr Trp Thr Ser Ser Phe Val Lys Phe Trp
    50                  55                  60
Glu Asn Met Asp Asp Thr Asn Ile Ile Ile Ile Tyr Gly Asp Ile Asp
65                  70                  75                  80
Ser Leu Glu Gly Leu Met Arg Asn Ile Gly Gln Arg Leu Leu Thr Trp
                85                  90                  95
His Val Trp Val Met Asn Ile Glu Pro His Ile Ile Glu Tyr Asp Asn
            100                 105                 110
Tyr Phe Met Leu Asp Ser Phe His Gly Ser Leu Ile Phe Lys His Asn
        115                 120                 125
Tyr Arg Glu Asn Phe Glu Phe Thr Lys Phe Ile Arg Thr Val Asn Pro
    130                 135                 140
Lys Lys Tyr Pro Glu Asp Ile Tyr Leu Pro Lys Met Trp Tyr Leu Phe
145                 150                 155                 160
```

```
Phe Met Cys Ser Phe Ser Asp Ile Asn Cys Gln Val Leu Asp Ser Cys
                165                 170                 175
Gln Thr Asn Ala Ser Leu Asp Met Leu Pro Ser Gln Ile Phe Asp Val
            180                 185                 190
Val Met Ser Glu Glu Ser Thr Ser Ile Tyr Asn Ala Val Tyr Ala Val
        195                 200                 205
Ala His Ser Leu His Glu Met Arg Leu Gln Gln Leu Gln Thr Gln Pro
    210                 215                 220
Cys Glu Asn Glu Glu Gly Met Glu Phe Phe Pro Trp Gln Leu Asn Thr
225                 230                 235                 240
Phe Leu Lys Asp Ile Glu Val Arg Val Asn Ser Leu Asp Trp Arg Gln
                245                 250                 255
Arg Ile Asp Ala Glu Tyr Asp Ile Leu Asn Leu Trp Asn Leu Pro Lys
            260                 265                 270
Gly Leu Gly Leu Lys Val Lys Ile Gly Asn Phe Tyr Ala Asn Ala Pro
        275                 280                 285
Gln Gly Gln Gln Leu Ser Leu Ser Glu Gln Met Ile Gln Trp Pro Glu
    290                 295                 300
Ile Phe Ser Glu Ile Pro Gln Ser Val Cys Ser Glu Ser Cys Gly Pro
305                 310                 315                 320
Gly Phe Arg Lys Val Thr Leu Glu Asn Lys Ala Ile Cys Cys Tyr Asn
                325                 330                 335
Cys Thr Pro Cys Ala Asp Asn Glu Ile Ser Asn Glu Thr Asp Val Asp
            340                 345                 350
Gln Cys Val Lys Cys Pro Glu Ser His Tyr Ala Asn Thr Glu Lys Ser
        355                 360                 365
Asn Cys Tyr Gln Lys Ser Val Ser Phe Leu Gly Tyr Glu Asp Pro Leu
    370                 375                 380
Gly Met Ala Leu Ala Ser Ile Ala Leu Cys Leu Ser Ala Leu Thr Ala
385                 390                 395                 400
Phe Val Ile Gly Ile Phe Val Lys His Lys Asp Thr Pro Ile Val Lys
                405                 410                 415
Ala Asn Asn Gln Ala Leu Ser Tyr Thr Leu Leu Ile Thr Leu Lys Phe
            420                 425                 430
Cys Phe Leu Cys Ser Leu Asn Phe Ile Gly Gln Pro Asn Thr Val Ala
        435                 440                 445
Cys Ile Leu Gln Gln Thr Thr Phe Ala Val Ala Phe Thr Met Ala Leu
    450                 455                 460
Ala Thr Val Leu Ala Lys Ala Ile Thr Val Val Leu Ala Phe Lys Val
465                 470                 475                 480
Ser Phe Pro Gly Arg Met Val Arg Trp Leu Met Ile Ser Arg Gly Pro
                485                 490                 495
Asn Tyr Ile Ile Pro Ile Cys Thr Leu Ile Gln Leu Leu Leu Cys Gly
            500                 505                 510
Ile Trp Met Ala Ile Ser Pro Pro Tyr Ile Asp Gln Asp Ala His Ile
        515                 520                 525
Glu His Gly His Ile Ile Ile Leu Cys Asn Lys Gly Ser Ala Val Ala
    530                 535                 540
Phe His Ser Val Leu Gly Tyr Leu Cys Phe Leu Ala Leu Gly Ser Tyr
545                 550                 555                 560
Thr Met Ala Phe Leu Ser Arg Asn Leu Pro Asp Thr Phe Asn Glu Ser
                565                 570                 575
Lys Phe Ile Ser Leu Ser Met Leu Val Phe Phe Cys Val Trp Ile Thr
            580                 585                 590
Phe Leu Pro Val Tyr His Ser Thr Lys Gly Lys Val
        595                 600
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1889 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
GAATTCGGCT TCTGCATCAA ATGGCGACGA AGGACACATC TCTTTCACTT GCCATTGTTT     60
CTTTGATGGT TCATTTTAGG TGGTCTTGGG TTGGTCTAAT TCTCCCAGAT GACCACAAAG    120
GAAATAAAAT ACTATCAGAT TTTAGAAAGG AGATGGAGAG AAAAAGAATC TGTACGGCTT    180
TTGTAAAAAT GATTCCTGCC ACATGGACTT CATCTTTTGT CAAATTCTGG GAAAATATGG    240
ATGACACCAA CATAATAATT ATTTATGGTG ACATTGATTC TCTAGAAGGT CCAATGCGAA    300
ATATTGGGCA AAGGTTATTG ACATGGCATG TCTGGGTCAT GAACATTGAA CCCCATATTA    360
TTGAATATGA TAATTATTTC ATGTTAGATT CATTCCATGG AAGTTTAATT TTTAAGCACA    420
ATTATAGAGA GAATTTTGAG TTTACCAAAT TTATTCGAAC AGTTAATCCT AAAAAATACC    480
CAGAAGACAT TTATCTCCCT AAGATGTGGT ATTTGTTCTT CATGTGCTCA TTTTCTGATA    540
TTAATTGTCA AGTTTTGGAC AGCTGTCAAA CAAATGCTTC TTTGGATATG TTACCTAGTC    600
AGATATTTGA TGTGGTCATG AGTGAAGAGA GCACAAGTAT TTACAATGCT GTGTACGCTG    660
TGGCTCACAG CCTCCATGAG ATGAGACTTC AGCAACTTCA AACACAACCG TGTGAAAATG    720
AAGAAGGGAT GGAGTTCTTT CCATGGCAGC TTAATACTTT CCTGAAGGAT ATTGAGGTGA    780
GAGTCAACAG TTTGGACTGG AGACAGAGAA TAGATGCTGA ATATGACATT CTTAACCTCT    840
GGAATTTACC AAAGGGTCTT GGACTAAAAG TGAAAATAGG AAACTTTTAT GCAAATGCTC    900
CCCAGGGTCA ACAATTGTCT TTATCTGAAC AGATGATTCA ATGGCCAGAA ATATTTTCAG    960
AAGTCCCTCA GTCTGTGTGC AGTGAGAGTT GTAGGCCTGG ATTCAGGAAA GTATCCCTGG   1020
ATGATAAGGC CATCTGCTGC TACAAGTGCA CTCCTTGTGC CGACAATGAG ATATCTAATG   1080
AGACAGATGT AGACCAGTGT GTGAAGTGTC CAGAGAGTCA TTATGCAAAT ACAGAGAAGA   1140
GCAACTGCTT CCCAAAATCT GTGAGCTTTC TGGCCTATGA AGACCCCTTG GGGATGGCTC   1200
TAGCCAGCAT AGCTTTGTGC TTATCTGCAC TCACTGTCTT TGTTATTGGC ATCTTTGTGA   1260
AAAACAGAGA CACTCCTATT GTCAAGGCCA ATAATCGGAC TCTAAGTTAC ATTTTGCTCA   1320
TCACACTCAC CTTTTGTTTC TTATGTTCTT TGAACTTCAT TGGTCAGCCC AACACAGCTG   1380
CCTGCATCCT TCAGCAGACC ACCTTTGCAG TTGCTTTCAC TATGGCTCTT GCCACTGTGT   1440
TGGCCAAAGC TATTACTGTA GTCCTTGCCT TTAAGATCAG TTTTCCAGGG AGAATGTTAA   1500
GGTGGCTAAT GATATCAAGG GGTCCAAGAT ACATCATTCC TATCTGCACA CTGATCCAGC   1560
TTCTTCTTTG TGGAATATGG ATGGCAACTT CTCCACCATT CATTGACCAA GATGTTAATA   1620
CTGAAGATGG ATACATCATC CTTTTGTGCA ACAAGGGCTC AGCTGTTGCC TTCCATTCAG   1680
TCCTGGGATA CCTCTGTTTC TTGGCCCTTG GGAGTTATAC CATGGCCTTC TTGTCTAGAA   1740
ATTTGCCTGA TACATTCAAT GAATCCAAAT TTCTGTCATT CAGTATGCTG GTGTTCTTCT   1800
GTGTCTGGGT CACCTTTCTT CCTGTCTACC ACAGCACTAA AGGGAAAGTT ATGGTCGTCG   1860
TCGAAGTCTT CTGCATCCAA GCCGAATTC                                     1889
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 604 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Ser Leu Ser Leu Ala Ile Val Ser Leu Met Val His Phe Arg Trp Ser
 1               5                  10                  15
Trp Val Gly Leu Ile Leu Pro Asp Asp His Lys Gly Asn Lys Ile Leu
            20                  25                  30
Ser Asp Phe Arg Lys Glu Met Glu Arg Lys Arg Ile Cys Thr Ala Phe
        35                  40                  45
Val Lys Met Ile Pro Ala Thr Trp Thr Ser Ser Phe Val Lys Phe Trp
    50                  55                  60
Glu Asn Met Asp Asp Thr Asn Ile Ile Ile Ile Tyr Gly Asp Ile Asp
65                  70                  75                  80
Ser Leu Glu Gly Pro Met Arg Asn Ile Gly Gln Arg Leu Leu Thr Trp
                85                  90                  95
His Val Trp Val Met Asn Ile Glu Pro His Ile Ile Glu Tyr Asp Asn
            100                 105                 110
```

Tyr Phe Met Leu Asp Ser Phe His Gly Ser Leu Ile Phe Lys His Asn
115 120 125
Tyr Arg Glu Asn Phe Glu Phe Thr Lys Phe Ile Arg Thr Val Asn Pro
130 135 140
Lys Lys Tyr Pro Glu Asp Ile Tyr Leu Pro Lys Met Trp Tyr Leu Phe
145 150 155 160
Phe Met Cys Ser Phe Ser Asp Ile Asn Cys Gln Val Leu Asp Ser Cys
165 170 175
Gln Thr Asn Ala Ser Leu Asp Met Leu Pro Ser Gln Ile Phe Asp Val
180 185 190
Val Met Ser Glu Glu Ser Thr Ser Ile Tyr Asn Ala Val Tyr Ala Val
195 200 205
Ala His Ser Leu His Glu Met Arg Leu Gln Gln Leu Gln Thr Gln Pro
210 215 220
Cys Glu Asn Glu Glu Gly Met Glu Phe Phe Pro Trp Gln Leu Asn Thr
225 230 235 240
Phe Leu Lys Asp Ile Glu Val Arg Val Asn Ser Leu Asp Trp Arg Gln
245 250 255
Arg Ile Asp Ala Glu Tyr Asp Ile Leu Asn Leu Trp Asn Leu Pro Lys
260 265 270
Gly Leu Gly Leu Lys Val Lys Ile Gly Asn Phe Tyr Ala Asn Ala Pro
275 280 285
Gln Gly Gln Gln Leu Ser Leu Ser Glu Gln Met Ile Gln Trp Pro Glu
290 295 300
Ile Phe Ser Glu Val Pro Gln Ser Val Cys Ser Glu Ser Cys Arg Pro
305 310 315 320
Gly Phe Arg Lys Val Ser Leu Asp Asp Lys Ala Ile Cys Cys Tyr Lys
325 330 335
Cys Thr Pro Cys Ala Asp Asn Glu Ile Ser Asn Glu Thr Asp Val Asp
340 345 350
Gln Cys Val Lys Cys Pro Glu Ser His Tyr Ala Asn Thr Glu Lys Ser
355 360 365
Asn Cys Phe Pro Lys Ser Val Ser Phe Leu Ala Tyr Glu Asp Pro Leu
370 375 380
Gly Met Ala Leu Ala Ser Ile Ala Leu Cys Leu Ser Ala Leu Thr Val
385 390 395 400
Phe Val Ile Gly Ile Phe Val Lys Asn Arg Asp Thr Pro Ile Val Lys
405 410 415
Ala Asn Asn Arg Thr Leu Ser Tyr Ile Leu Leu Ile Thr Leu Thr Phe
420 425 430
Cys Phe Leu Cys Ser Leu Asn Phe Ile Gly Gln Pro Asn Thr Ala Ala
435 440 445
Cys Ile Leu Gln Gln Thr Thr Phe Ala Val Ala Phe Thr Met Ala Leu
450 455 460
Ala Thr Val Leu Ala Lys Ala Ile Thr Val Val Leu Ala Phe Lys Ile
465 470 475 480
Ser Phe Pro Gly Arg Met Leu Arg Trp Leu Met Ile Ser Arg Gly Pro
485 490 495
Arg Tyr Ile Ile Pro Ile Cys Thr Leu Ile Gln Leu Leu Leu Cys Gly
500 505 510
Ile Trp Met Ala Thr Ser Pro Pro Phe Ile Asp Gln Asp Val Asn Thr
515 520 525
Glu Asp Gly Tyr Ile Ile Leu Leu Cys Asn Lys Gly Ser Ala Val Ala
530 535 540
Phe His Ser Val Leu Gly Tyr Leu Cys Phe Leu Ala Leu Gly Ser Tyr
545 550 555 560
Thr Met Ala Phe Leu Ser Arg Asn Leu Pro Asp Thr Phe Asn Glu Ser
565 570 575
Lys Phe Leu Ser Phe Ser Met Leu Val Phe Phe Cys Val Trp Val Thr
580 585 590
Phe Leu Pro Val Tyr His Ser Thr Lys Gly Lys Val
595 600

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2561 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 80...349
(D) OTHER INFORMATION: VR8

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
ATAGGTGCAA CTGTGTGTGT GATGTTTTTC TACATCAGAA ACGGATTTCA CAACAGCTCC     60
ATCTTAGATC CTAGCAGAC ATG AAG AAG CTC TGT GCT TTC ACG ATT TCA TTG     112
                     Met Lys Lys Leu Cys Ala Phe Thr Ile Ser Leu
                      1               5                  10

TTG TTT CTG AAG TTT TCT CTC ATC TTG TGC TGT TGG AGT GAA CCA AGT     160
Leu Phe Leu Lys Phe Ser Leu Ile Leu Cys Cys Trp Ser Glu Pro Ser
            15                  20                  25

TGC TTT TGG AGG ATA AAG AAT AGT GAT GAT AAT GAC GGA GAT TTG CAA     208
Cys Phe Trp Arg Ile Lys Asn Ser Asp Asp Asn Asp Gly Asp Leu Gln
        30                  35                  40

AGG GAA TGT CAT TTT TAC CTT GGG GCA GCT GAT ACA CCA GTT GAA GAT     256
Arg Glu Cys His Phe Tyr Leu Gly Ala Ala Asp Thr Pro Val Glu Asp
    45                  50                  55

AAT TTT TAT AGT TCA CTT TTA AAA TTT AGG TTT TCT TTG GAC CAT TTA     304
Asn Phe Tyr Ser Ser Leu Leu Lys Phe Arg Phe Ser Leu Asp His Leu
60                  65                  70                  75

ATC CTA ACC TAC GCG ACC ATG ACC GGC TGC CCC ATG TCC ATC AGG TAGCC   354
Ile Leu Thr Tyr Ala Thr Met Thr Gly Cys Pro Met Ser Ile Arg
                80                  85                  90

CCCAAGGACA CACATTTGTC CCATGGCATG GTCTCCTTGA TGTTTCACTT TAGATGGACT    414
TGGATAGGAA TGGTCATCTC AGATGATGAC CAGGGTATTC AGTTTCTCTC AGATTTAAGA    474
GAAGAAAGCC AAAGGCATGG GATCTGTTTA GCTTTTGTTA ATATGATCCC AGAAAACATG    534
CAGATATACA TGACAAGGGC TACAATATAT GATCAACAAA TTATGACATC TTCAGCAAAG    594
GTTGTTATCA TTTATGGTGA AATGAACTCT ACTCTAGAAG TAAGCTTTAG AAGATGGGAA    654
GAGTTAGGTG CTCGGAGAAT CTGGATCACA ACCTCACAAT GGGATGTCAT CACAAATAAA    714
AAAGACTTCA CCCTTAATCT CTTCCATGGG ACTATCACTT TTGCACACCA CAGAGTTGAG    774
ATTCCTAAAT TAAATAAATT CATGCAAACA ATGAACACTG CCAAATACCC AGTAGATATT    834
TCTCATACTA TATTGGAGTG GAATTATTTT AATTGTTCAA TATCTAAGAA CAGCATTAGA    894
ATGCATCATA TTACATTCAA CAACACCTTG GAATGGACAT CACTGCACAA CTATGATATG    954
GCGATGAGTG ATGAAGGTTA CAGTTTATAT AATGCTGTTT ATGCTGTGGC CCACACCTAC   1014
CATGAATACA TTTTTCAACA AGTAGAGTCT CAGAAAAAGG CAAAACCCAA AAGATATTTC   1074
ACTGCTTGTC AGCAGCCTCA GGTTCCCTCC TCCGTGTGTA GTGTGGCATG TACTGCTGGA   1134
TTCAGGAAAA TTTATCAAAA AGAAACAGCA GACTGCTGCT TTGATTGTGT TCAGTGCCCA   1194
GAAAATGAGA TTTCCAACGA AACAGATATG GAACAGTGTG TGAGGTGTCC AGATGATAAG   1254
TATGCCAACA TAGAGCAAAC CCACTGCCTC TCAAGAGCTG TATCATTTCT GGCTTATGAA   1314
GATCCATTGG GGATGGCTCT AGGCTGCATG GCACTGTCCT TCTCGGCCAT CACAATTCTA   1374
GTCCTCGTCA CATTTGTGAA ACACAACGAT ACTCCCATTG TGAAGGCCAA TAACCGCATT   1434
CTCAGCTACA TCCTGCTCAT CTCTCTCGTC TTCTGCTTTC TCTGCTCCCT GCTCTTCATT   1494
GGACCTCCCG ACCAGGTCAC CTGCATCTTG CAGCAGACCA CATTTGGAGT ATTTTTCACT   1554
GTGTCTGTTT CTACAGTGTT GGCCAAAACA ATAACTGTGG TCATGGCTTT CAAGCTCACT   1614
ACTCCAGGAA GAAGGATGAG AGGGATGATG ATGACAGGGG CACCTAAGTT GGTCATTCCC   1674
ATTTGTACCC TGATCCAACT TGTTCTCTGT GGAATCTGGT TGGTCACATC TCCTCCCTTT   1734
ATTGACAGAG ATATACAATC TGAGCATGGG AAGATTGTCA TTCTTTGCAA TAAAGGCTCA   1794
```

```
GTCATTGCCT TCCACGTCGT CCTGGGATAC TTGGGCTCCT TGGCTCTGGG GAGCTTCACT   1854
TTGGCTTTCT TGGCTAGGAA CCTTCCTGAC ACATTCAATG AAGCCAAGTT CCTAACTTTC   1914
AGCATGCTGG TGTTCTGCAG TGTCTGGATC ACCTTCCTCC CTGTCTACCA CAGCACCAGG   1974
GGGAGGGTCA TGGTGGTTGT GGAGGTTTTC TCCATCTTGG CTTCTAGTGC AGGGTTGCTA   2034
ATGTGTATCT TTGTCCCAAA GTGTTATGTT ATTTTAATTA GACCAGATTC AAATATTATA   2094
AAGAAACATA AAGGTAAAGT GCTTAATTGA AACTTTCATG GTATGAAAAT GTTAGATGAT   2154
ATTCAACTTA TCTTATTCTT CATCTTAATA AAAGCAGTAC TTCATCATAT AAAAAATAAA   2214
GTAATATACA GATTTATACT TACAAACTGG ACAGCAAACA TGAATATGTT GAGAACTGGG   2274
ATTCTCAATT GAGGAATGGC TACCAACATT TTGATCTGTG GTTTTGTGTT TAAGCCATGC   2334
ACTTAATTAA TGATTAACAT GAGGTTACCC TACTGTCTGT GAACAGCGCC ACCTCTAGGC   2394
ATGCTGTCCT TGAGTTATAA GAAAGGGTAC TGCATACACA ATGGACATGA AGCCAGTAAT   2454
CAACATTATT CCACTTGCTT TCATGGAGTT CTTACTTCCA AGTTCATGCC TTGACTTTAT   2514
TCAATGTTCT ATGACAAAGG TAGATAAATA AATAAACACT TTTCCTC                 2561
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 90 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
Met Lys Lys Leu Cys Ala Phe Thr Ile Ser Leu Leu Phe Leu Lys Phe
 1               5                  10                  15
Ser Leu Ile Leu Cys Cys Trp Ser Glu Pro Ser Cys Phe Trp Arg Ile
            20                  25                  30
Lys Asn Ser Asp Asp Asn Asp Gly Asp Leu Gln Arg Glu Cys His Phe
        35                  40                  45
Tyr Leu Gly Ala Ala Asp Thr Pro Val Glu Asp Asn Phe Tyr Ser Ser
    50                  55                  60
Leu Leu Lys Phe Arg Phe Ser Leu Asp His Leu Ile Leu Thr Tyr Ala
65                  70                  75                  80
Thr Met Thr Gly Cys Pro Met Ser Ile Arg
                85                  90
```

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2734 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 80...1387
(D) OTHER INFORMATION: VR9

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
ATAGGTGCAA CTGTGTGTGT GATGTTTTTC TACATCAGAA ACGGATTTCA CAACAGCTCC     60
ATCTTAGATC CTAGCAGAC ATG AAG AAG CTC TGT GCT TTC ACG ATT TCA TTG     112
                     Met Lys Lys Leu Cys Ala Phe Thr Ile Ser Leu
                      1               5                  10

TTG TTT CTG AAG TTT TCT CTC ATC TTG TGC TGT TGG AGT GAA CCA AGT     160
Leu Phe Leu Lys Phe Ser Leu Ile Leu Cys Cys Trp Ser Glu Pro Ser
```

```
                15                  20                  25

TGC TTT TGG AGG ATA AAG AAT AGT GAT GAT AAT GAC GGA GAT TTG CAA      208
Cys Phe Trp Arg Ile Lys Asn Ser Asp Asp Asn Asp Gly Asp Leu Gln
        30                  35                  40

AGG GAA TGT CAT TTT TAC CTT GGG GCA GCT GAT ACA CCA GTT GAA GAT      256
Arg Glu Cys His Phe Tyr Leu Gly Ala Ala Asp Thr Pro Val Glu Asp
    45                  50                  55

AAT TTT TAT AGT TCA CTT TTA AAA TTT AGA ATT GCA GCA AGT GAA TAT      304
Asn Phe Tyr Ser Ser Leu Leu Lys Phe Arg Ile Ala Ala Ser Glu Tyr
60                  65                  70                  75

GAG TTT CTT CTC GTA ATG TTT TTT GCT ATC GAT GAG ATC AAC AGG AAT      352
Glu Phe Leu Leu Val Met Phe Phe Ala Ile Asp Glu Ile Asn Arg Asn
                80                  85                  90

CCT TAT CTT TTA CCC AAC ATA ACT TTG ATG TTC TCC TTC ATT GGT GGA      400
Pro Tyr Leu Leu Pro Asn Ile Thr Leu Met Phe Ser Phe Ile Gly Gly
            95                  100                 105

AAC TGT CAG GAT TTA TTG AGA GTT ATG GAC CAA GCA TAT ACA CAA ATA      448
Asn Cys Gln Asp Leu Leu Arg Val Met Asp Gln Ala Tyr Thr Gln Ile
        110                 115                 120

AAT GGA CAT ATG AAT TTT GTT AAT TAT TTC TGT TAT TTA GAT GAT TCA      496
Asn Gly His Met Asn Phe Val Asn Tyr Phe Cys Tyr Leu Asp Asp Ser
    125                 130                 135

TGT GCC ATA GGT CTT ACA GGA CCA TCA TGG AAA ACT TCC TTA AAA CTG      544
Cys Ala Ile Gly Leu Thr Gly Pro Ser Trp Lys Thr Ser Leu Lys Leu
140                 145                 150                 155

GCA ATG CAC TCT TCG ATG CCA CTG GTT TTC TTT GGA CCA TTT AAT CCT      592
Ala Met His Ser Ser Met Pro Leu Val Phe Phe Gly Pro Phe Asn Pro
                160                 165                 170

AAC CTA CGC GAC CAT GAC CGG CTG CCC CAT GTC CAT CAG GTA GCC CCC      640
Asn Leu Arg Asp His Asp Arg Leu Pro His Val His Gln Val Ala Pro
            175                 180                 185

AAG GAC ACA CAT TTG TCC CAT GGC ATG GTC TCC TTG ATG TTT CAC TTT      688
Lys Asp Thr His Leu Ser His Gly Met Val Ser Leu Met Phe His Phe
        190                 195                 200

AGA TGG ACT TGG ATA GGA ATG GTC ATC TCA GAT GAT GAC CAG GGT ATT      736
Arg Trp Thr Trp Ile Gly Met Val Ile Ser Asp Asp Asp Gln Gly Ile
    205                 210                 215

CAG TTT CTC TCA GAT TTA AGA GAA GAA AGC CAA AGG CAT GGG ATC TGT      784
Gln Phe Leu Ser Asp Leu Arg Glu Glu Ser Gln Arg His Gly Ile Cys
220                 225                 230                 235

TTA GCT TTT GTT AAT ATG ATC CCA GAA AAC ATG CAG ATA TAC ATG ACA      832
Leu Ala Phe Val Asn Met Ile Pro Glu Asn Met Gln Ile Tyr Met Thr
                240                 245                 250

AGG GCT ACA ATA TAT GAT CAA CAA ATT ATG ACA TCT TCA GCA AAG GTT      880
Arg Ala Thr Ile Tyr Asp Gln Gln Ile Met Thr Ser Ser Ala Lys Val
            255                 260                 265

GTT ATC ATT TAT GGT GAA ATG AAC TCT ACT CTA GAA GTA AGC TTT AGA      928
Val Ile Ile Tyr Gly Glu Met Asn Ser Thr Leu Glu Val Ser Phe Arg
        270                 275                 280
```

```
AGA TGG GAA GAG TTA GGT GCT CGG AGA ATC TGG ATC ACA ACC TCA CAA      976
Arg Trp Glu Glu Leu Gly Ala Arg Arg Ile Trp Ile Thr Thr Ser Gln
    285                 290                 295

TGG GAT GTC ATC ACA AAT AAA AAA GAC TTC ACC CTT AAT CTC TTC CAT     1024
Trp Asp Val Ile Thr Asn Lys Lys Asp Phe Thr Leu Asn Leu Phe His
300                 305                 310                 315

GGG ACT ATC ACT TTT GCA CAC CAC AGA GTT GAG ATT CCT AAA TTA AAT     1072
Gly Thr Ile Thr Phe Ala His His Arg Val Glu Ile Pro Lys Leu Asn
                320                 325                 330

AAA TTC ATG CAA ACA ATG AAC ACT GCC AAA TAC CCA GTA GAT ATT TCT     1120
Lys Phe Met Gln Thr Met Asn Thr Ala Lys Tyr Pro Val Asp Ile Ser
            335                 340                 345

CAT ACT ATA TTG GAG TGG AAT TAT TTT AAT TGT TCA ATA TCT AAG AAC     1168
His Thr Ile Leu Glu Trp Asn Tyr Phe Asn Cys Ser Ile Ser Lys Asn
        350                 355                 360

AGC ATT AGA ATG CAT CAT ATT ACA TTC AAC AAC ACC TTG GAA TGG ACA     1216
Ser Ile Arg Met His His Ile Thr Phe Asn Asn Thr Leu Glu Trp Thr
    365                 370                 375

TCA CTG CAC AAC TAT GAT ATG GCG ATG AGT GAT GAA GGT TAC AGT TTA     1264
Ser Leu His Asn Tyr Asp Met Ala Met Ser Asp Glu Gly Tyr Ser Leu
380                 385                 390                 395

TAT AAT GCT GTT TAT GCT GTG GCC CAC ACC TAC CAT GAA TAC ATT TTT     1312
Tyr Asn Ala Val Tyr Ala Val Ala His Thr Tyr His Glu Tyr Ile Phe
                400                 405                 410

CAA CAA GTA GAG TCT CAG AAA AAG GCA AAA CCC AAA AGA TAT TTC ACT     1360
Gln Gln Val Glu Ser Gln Lys Lys Ala Lys Pro Lys Arg Tyr Phe Thr
            415                 420                 425

GCT TGT CAG CAG ATA TGG AAC AGT GTG TGAGGTGTCC AGATGATAAG TATGCCA   1414
Ala Cys Gln Gln Ile Trp Asn Ser Val
        430                 435

ACATAGAGCA AACCCACTGC CTCTCAAGAG CTGTATCATT TCTGGCTTAT GAAGATCCAT   1474
TGGGGATGGC TCTAGGCTGC ATGGCACTGT CCTTCTCGGC CATCACAATT CTAGTCCTCG   1534
TCACATTTGT GAAACACAAC GATACTCCCA TTGTGAAGGC CAATAACCGC ATTCTCAGCT   1594
ACATCCTGCT CATCTCTCTC GTCTTCTGCT TTCTCTGCTC CCTGCTCTTC ATTGGACCTC   1654
CCGACCAGGT CACCTGCATC TTGCAGCAGA CCACATTTGG AGTATTTTTC ACTGTGTCTG   1714
TTTCTACAGT GTTGGCCAAA ACAATAACTG TGGTCATGGC TTTCAAGCTC ACTACTCCAG   1774
GAAGAAGGAT GAGAGGGATG ATGATGACAG GGGCACCTAA GTTGGTCATT CCCATTTGTA   1834
CCCTGATCCA ACTTGTTCTC TGTGGAATCT GGTTGGTCAC ATCTCCTCCC TTTATTGACA   1894
GAGATATACA ATCTGAGCAT GGGAAGATTG TCATTCTTTG CAATAAAGGC TCAGTCATTG   1954
CCTTCCACGT CGTCCTGGGA TACTTGGGCT CCTTGGCTCT GGGGAGCTTC ACTTTGGCTT   2014
TCTTGGCTAG GAACCTTCCT GACACATTCA ATGAAGCCAA GTTCCTAACT TTCAGCATGC   2074
TGGTGTTCTG CAGTGTCTGG ATCACCTTCC TCCCTGTCTA CCACAGCACC AGGGGGAGGG   2134
TCATGGTGGT TGTGGAGGTT TTCTCCATCT TGGCTTCTAG TGCAGGGTTG CTAATGTGTA   2194
TCTTTGTCCC AAAGTGTTAT GTTATTTTAA TTAGACCAGA TTCAAATATT ATAAAGAAAC   2254
ATAAAGGTAA AGTGCTTAAT TGAAACTTTC ATGGTATGAA AATGTTAGAT GATATTCAAC   2314
TTATCTTATT CTTCATCTTA ATAAAAGCAG TACTTCATCA TATAAAAAAT AAAGTAATAT   2374
ACAGATTTAT ACTTACAAAC TGGACAGCAA ACATGAATAT GTTGAGAACT GGGATTCTCA   2434
ATTGAGGAAT GGCTACCAAC ATTTTGATCT GTGGTTTTGT GTTTAAGCCA TGCACTTAAT   2494
TAATGATTAA CATGAGGTTA CCCTACTGTC TGTGAACAGC GCCACCTCTA GGCATGCTGT   2554
CCTTGAGTTA TAAGAAAGGG TACTGCATAC ACAATGGACA TGAAGCCAGT AATCAACATT   2614
ATTCCACTTG CTTTCATGGA GTTCTTACTT CCAAGTTCAT GCCTTGACTT TATTCAATGT   2674
TCTATGACAA AGGTAGATAA ATAAATAAAC ACTTTCCTCA CAAAAAAAAA AAAAAAAAAA   2734
                                                                    2734
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 436 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
Met Lys Lys Leu Cys Ala Phe Thr Ile Ser Leu Leu Phe Leu Lys Phe
 1               5                  10                  15
Ser Leu Ile Leu Cys Cys Trp Ser Glu Pro Ser Cys Phe Trp Arg Ile
            20                  25                  30
Lys Asn Ser Asp Asp Asn Asp Gly Asp Leu Gln Arg Glu Cys His Phe
        35                  40                  45
Tyr Leu Gly Ala Ala Asp Thr Pro Val Glu Asp Asn Phe Tyr Ser Ser
    50                  55                  60
Leu Leu Lys Phe Arg Ile Ala Ala Ser Glu Tyr Glu Phe Leu Leu Val
65                  70                  75                  80
Met Phe Phe Ala Ile Asp Glu Ile Asn Arg Asn Pro Tyr Leu Leu Pro
                85                  90                  95
Asn Ile Thr Leu Met Phe Ser Phe Ile Gly Gly Asn Cys Gln Asp Leu
            100                 105                 110
Leu Arg Val Met Asp Gln Ala Tyr Thr Gln Ile Asn Gly His Met Asn
        115                 120                 125
Phe Val Asn Tyr Phe Cys Tyr Leu Asp Asp Ser Cys Ala Ile Gly Leu
    130                 135                 140
Thr Gly Pro Ser Trp Lys Thr Ser Leu Lys Leu Ala Met His Ser Ser
145                 150                 155                 160
Met Pro Leu Val Phe Phe Gly Pro Phe Asn Pro Asn Leu Arg Asp His
                165                 170                 175
Asp Arg Leu Pro His Val His Gln Val Ala Pro Lys Asp Thr His Leu
            180                 185                 190
Ser His Gly Met Val Ser Leu Met Phe His Phe Arg Trp Thr Trp Ile
        195                 200                 205
Gly Met Val Ile Ser Asp Asp Asp Gln Gly Ile Gln Phe Leu Ser Asp
    210                 215                 220
Leu Arg Glu Glu Ser Gln Arg His Gly Ile Cys Leu Ala Phe Val Asn
225                 230                 235                 240
Met Ile Pro Glu Asn Met Gln Ile Tyr Met Thr Arg Ala Thr Ile Tyr
                245                 250                 255
Asp Gln Gln Ile Met Thr Ser Ser Ala Lys Val Val Ile Ile Tyr Gly
            260                 265                 270
Glu Met Asn Ser Thr Leu Glu Val Ser Phe Arg Arg Trp Glu Glu Leu
        275                 280                 285
Gly Ala Arg Arg Ile Trp Ile Thr Thr Ser Gln Trp Asp Val Ile Thr
    290                 295                 300
Asn Lys Lys Asp Phe Thr Leu Asn Leu Phe His Gly Thr Ile Thr Phe
305                 310                 315                 320
Ala His His Arg Val Glu Ile Pro Lys Leu Asn Lys Phe Met Gln Thr
                325                 330                 335
Met Asn Thr Ala Lys Tyr Pro Val Asp Ile Ser His Thr Ile Leu Glu
            340                 345                 350
Trp Asn Tyr Phe Asn Cys Ser Ile Ser Lys Asn Ser Ile Arg Met His
        355                 360                 365
His Ile Thr Phe Asn Asn Thr Leu Glu Trp Thr Ser Leu His Asn Tyr
    370                 375                 380
Asp Met Ala Met Ser Asp Glu Gly Tyr Ser Leu Tyr Asn Ala Val Tyr
385                 390                 395                 400
Ala Val Ala His Thr Tyr His Glu Tyr Ile Phe Gln Gln Val Glu Ser
                405                 410                 415
Gln Lys Lys Ala Lys Pro Lys Arg Tyr Phe Thr Ala Cys Gln Gln Ile
            420                 425                 430
```

```
Trp Asn Ser Val
        435
```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2732 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 80...1375
(D) OTHER INFORMATION: VR10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```
ATAGTTGTAA ATGTGTGTGT GATGTTTTTC TACATCAGAA ACGGATTTCA CAACAACTCC     60
ATCTTAGATC CTAGCAGAC ATG AAG AAG CTC TGT GCT TTC ACT ATT TCA TTT     112
                     Met Lys Lys Leu Cys Ala Phe Thr Ile Ser Phe
                      1               5                  10

TTG TCT CTG AAG TTT TCT CTC ATC TTG TGC TGT TTG ACT GAA GCA AGT     160
Leu Ser Leu Lys Phe Ser Leu Ile Leu Cys Cys Leu Thr Glu Ala Ser
            15                  20                  25

TGC TTT TGG AGG ATA AAG AAT AGT GAA GAT AGT GAT GGA GAT TTG CAA     208
Cys Phe Trp Arg Ile Lys Asn Ser Glu Asp Ser Asp Gly Asp Leu Gln
        30                  35                  40

AGA GAA TGT CAT TTT TAC CTT TGG GTA ATT GAT AAA CCT ATT GAA GAT     256
Arg Glu Cys His Phe Tyr Leu Trp Val Ile Asp Lys Pro Ile Glu Asp
    45                  50                  55

AAT TTT TAT AAT TCA GTT TTA AAT TTT AGA ATA TCA GCA AGT GAA TAT     304
Asn Phe Tyr Asn Ser Val Leu Asn Phe Arg Ile Ser Ala Ser Glu Tyr
60                  65                  70                  75

GAG TTT CTT CTG GTA ATG TTT TTT GCT ACT GAT GAG ATC AAC AAG AAT     352
Glu Phe Leu Leu Val Met Phe Phe Ala Thr Asp Glu Ile Asn Lys Asn
                80                  85                  90

CCT TAT CTT TTA CCC AAC ATA ACT TTG ATA TTC AGC ATC GTT GGT GGT     400
Pro Tyr Leu Leu Pro Asn Ile Thr Leu Ile Phe Ser Ile Val Gly Gly
            95                  100                 105

CAC TGT CAT GAT TTA TTG AGA GGT CTG GAT CAA TCA TAT ACA CAA ATA     448
His Cys His Asp Leu Leu Arg Gly Leu Asp Gln Ser Tyr Thr Gln Ile
        110                 115                 120

AAT GGA CGT GTG AAT TTT GTT AAT TAT TTC TGT TAT TTA GAT GAT TCA     496
Asn Gly Arg Val Asn Phe Val Asn Tyr Phe Cys Tyr Leu Asp Asp Ser
    125                 130                 135

TGT AAC ATA GGC CTT ACA GGA CCA TCA TGG AAA AAA TCC TTA AAA CTG     544
Cys Asn Ile Gly Leu Thr Gly Pro Ser Trp Lys Lys Ser Leu Lys Leu
140                 145                 150                 155

GCA ATG GAT TCT TCA ATA CCA ATG GTT TTC TTT GGA CCA TTT AAT CCT     592
Ala Met Asp Ser Ser Ile Pro Met Val Phe Phe Gly Pro Phe Asn Pro
                160                 165                 170
```

```
AAC CTA CGC GAC CAT GAC CGG CTG CCC CAT GTC CAT CAG GTA GCC CCC      640
Asn Leu Arg Asp His Asp Arg Leu Pro His Val His Gln Val Ala Pro
            175                 180                 185

AAG GAC ACA CAT TTA TCC CAT GGC ATG GTC TCC TTG ATG TTT CAT TTT      688
Lys Asp Thr His Leu Ser His Gly Met Val Ser Leu Met Phe His Phe
        190                 195                 200

AGA TGG ACT TGG ATA GGA CTG GTC ATC TCA GAT GAT GAC CAG GGT ATT      736
Arg Trp Thr Trp Ile Gly Leu Val Ile Ser Asp Asp Asp Gln Gly Ile
    205                 210                 215

CAG TTT CTC TCA GAT TTA AGA GAA GAA AGC CAA AGG CAT GGG ATC TGT      784
Gln Phe Leu Ser Asp Leu Arg Glu Glu Ser Gln Arg His Gly Ile Cys
220                 225                 230                 235

TTA GCT TTT GTT AAT ATG ATC CCA GAA AAC ATG CAG ATA TAC ATG ACA      832
Leu Ala Phe Val Asn Met Ile Pro Glu Asn Met Gln Ile Tyr Met Thr
                240                 245                 250

AGG GCT ACA ATA TAT GAT AAA CAA ATT ATG ACA TCT TCA GCA AAG GTT      880
Arg Ala Thr Ile Tyr Asp Lys Gln Ile Met Thr Ser Ser Ala Lys Val
            255                 260                 265

GTT ATC ATT TAT GGT GAA ATG AAC TCT ACT CTA GAA GTA AGC TTC AGA      928
Val Ile Ile Tyr Gly Glu Met Asn Ser Thr Leu Glu Val Ser Phe Arg
        270                 275                 280

AGA TGG GAA GAT TTA GGT GCT CGG AGA ATC TGG ATC ACA ACC TCA CAA      976
Arg Trp Glu Asp Leu Gly Ala Arg Arg Ile Trp Ile Thr Thr Ser Gln
    285                 290                 295

TGG GAT ATC ATA TTA AAT AAA AAA GAA TTC ACT CTT AAT CTC TTC CAT     1024
Trp Asp Ile Ile Leu Asn Lys Lys Glu Phe Thr Leu Asn Leu Phe His
300                 305                 310                 315

GGC CCT ATC ACT TTT GCA CAC CAC AAA GTT GAG ATT CCT AAA TTA AGG     1072
Gly Pro Ile Thr Phe Ala His His Lys Val Glu Ile Pro Lys Leu Arg
                320                 325                 330

AAT TTT ATG CAA ACA ATG AAC ACT GCC AAA TAC CCA GTA GAT ATT TCT     1120
Asn Phe Met Gln Thr Met Asn Thr Ala Lys Tyr Pro Val Asp Ile Ser
            335                 340                 345

CAT ACT ATA CTG GAG TGG AAT TAT TTT AAT TGT TCA ATC TCT AAG AAC     1168
His Thr Ile Leu Glu Trp Asn Tyr Phe Asn Cys Ser Ile Ser Lys Asn
        350                 355                 360

AGC AGT AAA ATG GAT CTT TTT ACA TCC AAC AAC ACA TTG GAA TGG ACA     1216
Ser Ser Lys Met Asp Leu Phe Thr Ser Asn Asn Thr Leu Glu Trp Thr
    365                 370                 375

GCA CTG CAC AAC TAT GAT ATG GCC ATG AGT GAT GAA GGT TAC AAT TTG     1264
Ala Leu His Asn Tyr Asp Met Ala Met Ser Asp Glu Gly Tyr Asn Leu
380                 385                 390                 395

TAT AAT GCT GTT TAT GTT GCG GCC CAC ACC TAC CAT GAA CAC ATT CTT     1312
Tyr Asn Ala Val Tyr Val Ala Ala His Thr Tyr His Glu His Ile Leu
                400                 405                 410

CAA CAA GTA GAG TCT CAG AAA AAG GTA GAA CAC AAC AGA TAT TTC ACT     1360
Gln Gln Val Glu Ser Gln Lys Lys Val Glu His Asn Arg Tyr Phe Thr
            415                 420                 425

GTT TGT CAG CAG ATA TAGAACAGTG TGTGAAATGT CCAGATGATA AGTATGCCAA C   1416
```

Val Cys Gln Gln Ile
430

ATAGAACAAA CCTACTGCCT CTCAAGAGCT GTATCATTTC TGGCTTTTGA AGAACCACTG 1476
GGGATGGCTC TAGGCTGCAT GGCACTATCC TTCTCGGCCA TCACAATTCT AGTACTAGTC 1536
ACATTTGTGA AGTACAAGAA TACTCCCATT GTGAAGGCCA ATAACCGCAT TCTCAGCTAC 1596
ATCCTGCTCA TCTCTCTAGT CTTCTGTTTT CTCTGCTCCC TGCTCTTCAT TGGACATCCT 1656
GACCAGGTCA CCTGCATCTT GCAGCAGACC ACATTTGGAG TATTTTTCAC TGTGTCTGTT 1716
TCTACAGTGT TGGCCAAAAC AATAACTGTG GTCATGGCTT TCAAGTTCAC TACTCCAGGA 1776
AGAAGGATGA GAGGGATGTT GGTAACAGGT GCACCTAAGT TGGTCATTCC CATTTGTACC 1836
CTAATCCAAC TTGTTCTCTG TGGAATCTGG TTGGTAACAT CTCCTCCATT TATTGACAGA 1896
GATATACAAT CTGAACATGG GAAGGTAGTC ATTCTTTGCA ATAAAGGCTC TGTCATTGCC 1956
TTCCACATTG TCCTGGGATA CTTGGGCTCC TTGGCTCTGG GGAGCTTCAC TTTGGCTTTC 2016
TTGGCTAGGA ACCTTCCTGA CACATTCAAT GAAGCCAAAT TCCTAACTTT CAGCATGCTG 2076
GTGTTCTGCA GTGTCTGGAT CACCTTCCTC CCTGTCTACC ACAGCACCAG GGGGAAGGTC 2136
ATGGTGGTTG TGGAGGTTTT CTCAATCTTG GCTTCTAGTG CAGGGTTGCT AATGTGTATC 2196
TTTGTCCCAA AGTGTTATGT TATTTTAGTT AGACCAGATT CAAATTTTAC AAAGAACCGC 2256
AAAGGTAAAT TGCTTTATTG AAATTTTCAT GGTATGAAAA TGTTAGATTA TATTCAACTT 2316
ATCTTATTCT TCATCTTAAC AAAAGTAGTA CTTCATCATA TAAAAAATTA AGTAATATAC 2376
AGATTTATAC TTACAAACTG GACAGCAAAC ATGAATATGT TTAGAACTGG GAATCTCAAT 2436
TGAGGAATGG GTATCATCAT TTTGACCTGT GGTTATGTGT TTAAGCCATG TGTTTAATTA 2496
ATGATTAACA TGAGGTTGCC CTACTGTCTG TGAACCATAC CACCTCTAGG CACACTGTCC 2556
TTGAGTTATA AGATAGGGTA CTGCATACAA AATGGACATG AAACCAGTAA TCAACATTAT 2616
CCCTCTTGCT TTCATGGAGT TCTTGCATCC AATTTCATGC CTTGACTTCA TTCAATGTAC 2676
TATGACAAAG GTACATAAAT AAATAAACAC TTTCCCCACC AAAAAAAAAA AAAAAA 2732

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 432 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Lys Lys Leu Cys Ala Phe Thr Ile Ser Phe Leu Ser Leu Lys Phe
1 5 10 15
Ser Leu Ile Leu Cys Cys Leu Thr Glu Ala Ser Cys Phe Trp Arg Ile
20 25 30
Lys Asn Ser Glu Asp Ser Asp Gly Asp Leu Gln Arg Glu Cys His Phe
35 40 45
Tyr Leu Trp Val Ile Asp Lys Pro Ile Glu Asp Asn Phe Tyr Asn Ser
50 55 60
Val Leu Asn Phe Arg Ile Ser Ala Ser Glu Tyr Glu Phe Leu Leu Val
65 70 75 80
Met Phe Phe Ala Thr Asp Glu Ile Asn Lys Asn Pro Tyr Leu Leu Pro
85 90 95
Asn Ile Thr Leu Ile Phe Ser Ile Val Gly Gly His Cys His Asp Leu
100 105 110
Leu Arg Gly Leu Asp Gln Ser Tyr Thr Gln Ile Asn Gly Arg Val Asn
115 120 125
Phe Val Asn Tyr Phe Cys Tyr Leu Asp Asp Ser Cys Asn Ile Gly Leu
130 135 140
Thr Gly Pro Ser Trp Lys Lys Ser Leu Lys Leu Ala Met Asp Ser Ser
145 150 155 160
Ile Pro Met Val Phe Phe Gly Pro Phe Asn Pro Asn Leu Arg Asp His
165 170 175
Asp Arg Leu Pro His Val His Gln Val Ala Pro Lys Asp Thr His Leu
180 185 190
Ser His Gly Met Val Ser Leu Met Phe His Phe Arg Trp Thr Trp Ile
195 200 205

```
Gly Leu Val Ile Ser Asp Asp Asp Gln Gly Ile Gln Phe Leu Ser Asp
    210                 215                 220
Leu Arg Glu Glu Ser Gln Arg His Gly Ile Cys Leu Ala Phe Val Asn
225                 230                 235                 240
Met Ile Pro Glu Asn Met Gln Ile Tyr Met Thr Arg Ala Thr Ile Tyr
                245                 250                 255
Asp Lys Gln Ile Met Thr Ser Ser Ala Lys Val Val Ile Ile Tyr Gly
            260                 265                 270
Glu Met Asn Ser Thr Leu Glu Val Ser Phe Arg Arg Trp Glu Asp Leu
        275                 280                 285
Gly Ala Arg Arg Ile Trp Ile Thr Thr Ser Gln Trp Asp Ile Ile Leu
    290                 295                 300
Asn Lys Lys Glu Phe Thr Leu Asn Leu Phe His Gly Pro Ile Thr Phe
305                 310                 315                 320
Ala His His Lys Val Glu Ile Pro Lys Leu Arg Asn Phe Met Gln Thr
                325                 330                 335
Met Asn Thr Ala Lys Tyr Pro Val Asp Ile Ser His Thr Ile Leu Glu
            340                 345                 350
Trp Asn Tyr Phe Asn Cys Ser Ile Ser Lys Asn Ser Ser Lys Met Asp
        355                 360                 365
Leu Phe Thr Ser Asn Asn Thr Leu Glu Trp Thr Ala Leu His Asn Tyr
    370                 375                 380
Asp Met Ala Met Ser Asp Glu Gly Tyr Asn Leu Tyr Asn Ala Val Tyr
385                 390                 395                 400
Val Ala Ala His Thr Tyr His Glu His Ile Leu Gln Gln Val Glu Ser
                405                 410                 415
Gln Lys Lys Val Glu His Asn Arg Tyr Phe Thr Val Cys Gln Gln Ile
            420                 425                 430
```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2962 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 81...1601
(D) OTHER INFORMATION: VR11

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```
CATAGTTGTA AATGTGTGTG TGATGTTTTT CTACATCAGA AACGGATTTC ACAACAACTC     60
CATCTTAGAT CCTAGCAGAC  ATG AAG AAG CTC TGT GCT TTC ACT ATT TCA      110
                       Met Lys Lys Leu Cys Ala Phe Thr Ile Ser
                        1               5                  10

TTT TTG TCT CTG AAG TTT TCT CTC ATC TTG TGC TGT TTG ACT GAA GCA     158
Phe Leu Ser Leu Lys Phe Ser Leu Ile Leu Cys Cys Leu Thr Glu Ala
                15                  20                  25

AGT TGC TTT TGG AGG ATA AAG AAT AGT GAA GAT AGT GAT GGA GAT TTG     206
Ser Cys Phe Trp Arg Ile Lys Asn Ser Glu Asp Ser Asp Gly Asp Leu
            30                  35                  40

CAA AGA GAA TGT CAT TTT TAC CTT TGG GTA ATT GAT AAA CCT ATT GAA     254
Gln Arg Glu Cys His Phe Tyr Leu Trp Val Ile Asp Lys Pro Ile Glu
        45                  50                  55

GAT AAT TTT TAT AAT TCA GTT TTA AAT TTT AGA ATA TCA GCA AGT GAA     302
```

```
Asp Asn Phe Tyr Asn Ser Val Leu Asn Phe Arg Ile Ser Ala Ser Glu
    60                  65                  70

TAT GAG TTT CTT CTG GTA ATG TTT TTT GCT ACT GAT GAG ATC AAC AAG      350
Tyr Glu Phe Leu Leu Val Met Phe Phe Ala Thr Asp Glu Ile Asn Lys
75                  80                  85                  90

AAT CCT TAT CTT TTA CCC AAC ATA ACT TTG ATA TTC AGC ATC GTT GGT      398
Asn Pro Tyr Leu Leu Pro Asn Ile Thr Leu Ile Phe Ser Ile Val Gly
                95                 100                 105

GGT CAC TGT CAT GAT TTA TTG AGA GGT CTG GAT CAA TCA TAT ACA CAA      446
Gly His Cys His Asp Leu Leu Arg Gly Leu Asp Gln Ser Tyr Thr Gln
            110                 115                 120

ATA AAT GGA CGT GTG AAT TTT GTT AAT TAT TTC TGT TAT TTA GAT GAT      494
Ile Asn Gly Arg Val Asn Phe Val Asn Tyr Phe Cys Tyr Leu Asp Asp
        125                 130                 135

TCA TGT AAC ATA GGC CTT ACA GGA CCA TCA TGG AAA AAA TCC TTA AAA      542
Ser Cys Asn Ile Gly Leu Thr Gly Pro Ser Trp Lys Lys Ser Leu Lys
    140                 145                 150

CTG GCA ATG GAT TCT TCA ATA CCA ATG GTT TTC TTT GGA CCA TTT AAT      590
Leu Ala Met Asp Ser Ser Ile Pro Met Val Phe Phe Gly Pro Phe Asn
155                 160                 165                 170

CCT AAC CTA CGC GAC CAT GAC CGG CTG CCC CAT GTC CAT CAG GTA GCC      638
Pro Asn Leu Arg Asp His Asp Arg Leu Pro His Val His Gln Val Ala
                175                 180                 185

CCC AAG GAC ACA CAT TTA TCC CAT GGC ATG GTC TCC TTG ATG TTT CAT      686
Pro Lys Asp Thr His Leu Ser His Gly Met Val Ser Leu Met Phe His
            190                 195                 200

TTT AGA TGG ACT TGG ATA GGA CTG GTC ATC TCA GAT GAT GAC CAG GGT      734
Phe Arg Trp Thr Trp Ile Gly Leu Val Ile Ser Asp Asp Asp Gln Gly
        205                 210                 215

ATT CAG TTT CTC TCA GAT TTA AGA GAA GAA AGC CAA AGG CAT GGG ATC      782
Ile Gln Phe Leu Ser Asp Leu Arg Glu Glu Ser Gln Arg His Gly Ile
    220                 225                 230

TGT TTA GCT TTT GTT AAT ATG ATC CCA GAA AAC ATG CAG ATA TAC ATG      830
Cys Leu Ala Phe Val Asn Met Ile Pro Glu Asn Met Gln Ile Tyr Met
235                 240                 245                 250

ACA AGG GCT ACA ATA TAT GAT AAA CAA ATT ATG ACA TCT TCA GCA AAG      878
Thr Arg Ala Thr Ile Tyr Asp Lys Gln Ile Met Thr Ser Ser Ala Lys
                255                 260                 265

GTT GTT ATC ATT TAT GGT GAA ATG AAC TCT ACT CTA GAA GTA AGC TTC      926
Val Val Ile Ile Tyr Gly Glu Met Asn Ser Thr Leu Glu Val Ser Phe
            270                 275                 280

AGA AGA TGG GAA GAT TTA GGT GCT CGG AGA ATC TGG ATC ACA ACC TCA      974
Arg Arg Trp Glu Asp Leu Gly Ala Arg Arg Ile Trp Ile Thr Thr Ser
        285                 290                 295

CAA TGG GAT ATC ATA TTA AAT AAA AAA GAA TTC ACT CTT AAT CTC TTC     1022
Gln Trp Asp Ile Ile Leu Asn Lys Lys Glu Phe Thr Leu Asn Leu Phe
    300                 305                 310

CAT GGC CCT ATC ACT TTT GCA CAC CAC AAA GTT GAG ATT CCT AAA TTA     1070
His Gly Pro Ile Thr Phe Ala His His Lys Val Glu Ile Pro Lys Leu
```

```
315                 320                 325                 330

AGG AAT TTT ATG CAA ACA ATG AAC ACT GCC AAA TAC CCA GTA GAT ATT    1118
Arg Asn Phe Met Gln Thr Met Asn Thr Ala Lys Tyr Pro Val Asp Ile
                335                 340                 345

TCT CAT ACT ATA CTG GAG TGG AAT TAT TTT AAT TGT TCA ATC TCT AAG    1166
Ser His Thr Ile Leu Glu Trp Asn Tyr Phe Asn Cys Ser Ile Ser Lys
            350                 355                 360

AAC AGC AGT AAA ATG GAT CTT TTT ACA TCC AAC AAC ACA TTG GAA TGG    1214
Asn Ser Ser Lys Met Asp Leu Phe Thr Ser Asn Asn Thr Leu Glu Trp
        365                 370                 375

ACA GCA CTG CAC AAC TAT GAT ATG GCC ATG AGT GAT GAA GGT TAC AAT    1262
Thr Ala Leu His Asn Tyr Asp Met Ala Met Ser Asp Glu Gly Tyr Asn
    380                 385                 390

TTG TAT AAT GCT GTT TAT GTT GCG GCC CAC ACC TAC CAT GAA CAC ATT    1310
Leu Tyr Asn Ala Val Tyr Val Ala Ala His Thr Tyr His Glu His Ile
395                 400                 405                 410

CTT CAA CAA GTA GAG TCT CAG AAA AAG GTA GAA CAC AAC AGA TAT TTC    1358
Leu Gln Gln Val Glu Ser Gln Lys Lys Val Glu His Asn Arg Tyr Phe
                415                 420                 425

ACT GTT TGT CAG CAG GTA TCT TCC TTG ATG AAA ACC AGG GTA TTT ACG    1406
Thr Val Cys Gln Gln Val Ser Ser Leu Met Lys Thr Arg Val Phe Thr
            430                 435                 440

AAC CCG GTT GGA GAA CTG GTG AAC ATG AAG CAT AGG GAA AAT CAG TGT    1454
Asn Pro Val Gly Glu Leu Val Asn Met Lys His Arg Glu Asn Gln Cys
        445                 450                 455

ACA GAG TAT GAT ATT TTC ATC ATT TGG AAT TTT CCA CAA GGC CTT GGA    1502
Thr Glu Tyr Asp Ile Phe Ile Ile Trp Asn Phe Pro Gln Gly Leu Gly
    460                 465                 470

TTA AAA TTG AAA ATA GGA AGC TAT ATA CCT TGT TTT CCA AAG AGT CAA    1550
Leu Lys Leu Lys Ile Gly Ser Tyr Ile Pro Cys Phe Pro Lys Ser Gln
475                 480                 485                 490

CAA CTT CAT ATA TCT GAT GAT TTG GAA TGG GCC ATG GGA GGA ACA TCA    1598
Gln Leu His Ile Ser Asp Asp Leu Glu Trp Ala Met Gly Gly Thr Ser
                495                 500                 505

ATA TAGAACAGTG TGTGAAATGT CCAGATGATA AGTATGCCAA CATAGAACAA ACCTAC  1657
Ile

TGCCTCTCAA GAGCTGTATC ATTTCTGGCT TTTGAAGAAC CACTGGGGAT GGCTCTAGGC  1717
TGCATGGCAC TATCCTTCTC GGCCATCACA ATTCTAGTAC TAGTCACATT TGTGAAGTAC  1777
AAGAATACTC CCATTGTGAA GGCCAATAAC CGCATTCTCA GCTACATCCT GCTCATCTCT  1837
CTAGTCTTCT GTTTTCTCTG CTCCCTGCTC TTCATTGGAC ATCCTGACCA GGTCACCTGC  1897
ATCTTGCAGC AGACCACATT TGGAGTATTT TTCACTGTGT CTGTTTCTAC AGTGTTGGCC  1957
AAAACAATAA CTGTGGTCAT GGCTTTCAAG TTCACTACTC CAGGAAGAAG GATGAGAGGG  2017
ATGTTGGTAA CAGGTGCACC TAAGTTGGTC ATTCCCATTT GTACCCTAAT CCAACTTGTT  2077
CTCTGTGGAA TCTGGTTGGT AACATCTCCT CCATTTATTG ACAGAGATAT ACAATCTGAA  2137
CATGGGAAGG TAGTCATTCT TTGCAATAAA GGCTCTGTCA TTGCCTTCCA CATTGTCCTG  2197
GGATACTTGG GCTCCTTGGC TCTGGGGAGC TTCACTTTGG CTTTCTTGGC TAGGAACCTT  2257
CCTGACACAT TCAATGAAGC CAAATTCCTA ACTTTCAGCA TGCTGGTGTT CTGCAGTGTC  2317
TGGATCACCT TCCTCCCTGT CTACCACAGC ACCAGGGGGA AGGTCATGGT GGTTGTGGAG  2377
GTTTTCTCAA TCTTGGCTTC TAGTGCAGGG TTGCTAATGT GTATCTTTGT CCCAAAGTGT  2437
TATGTTATTT TAGTTAGACC AGATTCAAAT TTTACAAAGA ACCGCAAAGG TAAATTGCTT  2497
TATTGAAATT TTCATGGTAT GAAAATGTTA GATTATATTC AACTTATCTT ATTCTTCATC  2557
```

```
TTAACAAAAG TAGTACTTCA TCATATAAAA AATTAAGTAA TATACAGATT TATACTTACA   2617
AACTGGACAG CAAACATGAA TATGTTTAGA ACTGGGAATC TCAATTGAGG AATGGGTATC   2677
ATCATTTTGA CCTGTGGTTA TGTGTTTAAG CCATGTGTTT AATTAATGAT TAACATGAGG   2737
TTGCCCTACT GTCTGTGAAC CATACCACCT CTAGGCACAC TGTCCTTGAG TTATAAGATA   2797
GGGTACTGCA TACAAAATGG ACATGAAACC AGTAATCAAC ATTATCCCTC TTGCTTTCAT   2857
GGAGTTCTTG CATCCAATTT CATGCCTTGA CTTCATTCAA TGTACTATGA CAAAGGTACA   2917
TAAATAAATA AACACTTTCC CCACAAAAAA AAAAAAAAAA AAAAA                   2962
```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 507 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
Met Lys Lys Leu Cys Ala Phe Thr Ile Ser Phe Leu Ser Leu Lys Phe
 1               5                  10                  15
Ser Leu Ile Leu Cys Cys Leu Thr Glu Ala Ser Cys Phe Trp Arg Ile
            20                  25                  30
Lys Asn Ser Glu Asp Ser Asp Gly Asp Leu Gln Arg Glu Cys His Phe
        35                  40                  45
Tyr Leu Trp Val Ile Asp Lys Pro Ile Glu Asp Asn Phe Tyr Asn Ser
    50                  55                  60
Val Leu Asn Phe Arg Ile Ser Ala Ser Glu Tyr Glu Phe Leu Leu Val
65                  70                  75                  80
Met Phe Phe Ala Thr Asp Glu Ile Asn Lys Asn Pro Tyr Leu Leu Pro
                85                  90                  95
Asn Ile Thr Leu Ile Phe Ser Ile Val Gly Gly His Cys His Asp Leu
            100                 105                 110
Leu Arg Gly Leu Asp Gln Ser Tyr Thr Gln Ile Asn Gly Arg Val Asn
        115                 120                 125
Phe Val Asn Tyr Phe Cys Tyr Leu Asp Asp Ser Cys Asn Ile Gly Leu
    130                 135                 140
Thr Gly Pro Ser Trp Lys Lys Ser Leu Lys Leu Ala Met Asp Ser Ser
145                 150                 155                 160
Ile Pro Met Val Phe Phe Gly Pro Phe Asn Pro Asn Leu Arg Asp His
                165                 170                 175
Asp Arg Leu Pro His Val His Gln Val Ala Pro Lys Asp Thr His Leu
            180                 185                 190
Ser His Gly Met Val Ser Leu Met Phe His Phe Arg Trp Thr Trp Ile
        195                 200                 205
Gly Leu Val Ile Ser Asp Asp Asp Gln Gly Ile Gln Phe Leu Ser Asp
    210                 215                 220
Leu Arg Glu Glu Ser Gln Arg His Gly Ile Cys Leu Ala Phe Val Asn
225                 230                 235                 240
Met Ile Pro Glu Asn Met Gln Ile Tyr Met Thr Arg Ala Thr Ile Tyr
                245                 250                 255
Asp Lys Gln Ile Met Thr Ser Ser Ala Lys Val Val Ile Ile Tyr Gly
            260                 265                 270
Glu Met Asn Ser Thr Leu Glu Val Ser Phe Arg Arg Trp Glu Asp Leu
        275                 280                 285
Gly Ala Arg Arg Ile Trp Ile Thr Thr Ser Gln Trp Asp Ile Ile Leu
    290                 295                 300
Asn Lys Lys Glu Phe Thr Leu Asn Leu Phe His Gly Pro Ile Thr Phe
305                 310                 315                 320
Ala His His Lys Val Glu Ile Pro Lys Leu Arg Asn Phe Met Gln Thr
                325                 330                 335
Met Asn Thr Ala Lys Tyr Pro Val Asp Ile Ser His Thr Ile Leu Glu
            340                 345                 350
```

```
Trp Asn Tyr Phe Asn Cys Ser Ile Ser Lys Asn Ser Ser Lys Met Asp
        355                 360                 365
Leu Phe Thr Ser Asn Asn Thr Leu Glu Trp Thr Ala Leu His Asn Tyr
    370                 375                 380
Asp Met Ala Met Ser Asp Glu Gly Tyr Asn Leu Tyr Asn Ala Val Tyr
385                 390                 395                 400
Val Ala Ala His Thr Tyr His Glu His Ile Leu Gln Gln Val Glu Ser
                405                 410                 415
Gln Lys Lys Val Glu His Asn Arg Tyr Phe Thr Val Cys Gln Gln Val
            420                 425                 430
Ser Ser Leu Met Lys Thr Arg Val Phe Thr Asn Pro Val Gly Glu Leu
        435                 440                 445
Val Asn Met Lys His Arg Glu Asn Gln Cys Thr Glu Tyr Asp Ile Phe
    450                 455                 460
Ile Ile Trp Asn Phe Pro Gln Gly Leu Gly Leu Lys Leu Lys Ile Gly
465                 470                 475                 480
Ser Tyr Ile Pro Cys Phe Pro Lys Ser Gln Gln Leu His Ile Ser Asp
                485                 490                 495
Asp Leu Glu Trp Ala Met Gly Gly Thr Ser Ile
            500                 505
```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2821 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 60...992
(D) OTHER INFORMATION: VR12

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```
GACGTTTTTC TGCATCAGAA ACGGATTTCA CAGCAGCTCC ATCTCAGATC CTAGCAGAC A     60
                                                                   Me

TGA AGC AGC TCT GCA CTT TCA CTA TTT CAT TGT TGT TTC TGA AGT TTT      108
 t Lys Gln Leu Cys Thr Phe Thr Ile Ser Leu Leu Phe Leu Lys Phe Se
1               5                  10                  15

CTC TCA TCT TGT GCT GTT GGA GTG AAC CAA GCT GCT TTT GGA GGA TAA      156
r Leu Ile Leu Cys Cys Trp Ser Glu Pro Ser Cys Phe Trp Arg Ile Ly
        20                  25                  30

AGA AGA GTG AAG ATA ATG ATG GAG ATT TAC AAA GGG AGT GTC ATT TTT      204
s Lys Ser Glu Asp Asn Asp Gly Asp Leu Gln Arg Glu Cys His Phe Ty
      35                  40                  45

ACC TTT GGA AAA CTG ATG AAC CTA TTG AAG ATA GTT TTT ATA ATT ATG      252
r Leu Trp Lys Thr Asp Glu Pro Ile Glu Asp Ser Phe Tyr Asn Tyr As
  50                  55                  60                   6

ATT TAA GTT TTA GAA TTG CAG GAA GTG AAT ATG AGC TTC TTC TGG TAA      300
 p Leu Ser Phe Arg Ile Ala Gly Ser Glu Tyr Glu Leu Leu Leu Val Me
5                   70                  75                  80

TGT TTT TTG CTA CTG ATG AGA TCA ACA AGA ATC CTT ATC TTT TAC CCA      348
 t Phe Phe Ala Thr Asp Glu Ile Asn Lys Asn Pro Tyr Leu Leu Pro As
```

```
                85                  90                  95

ACA TGA GTT TGA TGT TCT CCA TCA TTG GTG GAA ACT GTC ATG ATT TAT      396
n Met Ser Leu Met Phe Ser Ile Ile Gly Gly Asn Cys His Asp Leu Le
          100                 105                 110

TGA GAA GTC TGG ATC AAG AAT ATG CAC AAA TAG ATG GAC ATA TGA ATT      444
u Arg Ser Leu Asp Gln Glu Tyr Ala Gln Ile Asp Gly His Met Asn Ph
      115                 120                 125

TTG TTA ATT ATT TCT GTT ATT TAG ATG ATT CAT GTG CCA CAG GCC TTA      492
e Val Asn Tyr Phe Cys Tyr Leu Asp Asp Ser Cys Ala Thr Gly Leu Th
  130                 135                 140                 1

CAG GAC CAT CAT GGA AAA CAT CCT TAA AAC TGG CAA TGC ATT CTT CAA      540
r Gly Pro Ser Trp Lys Thr Ser Leu Lys Leu Ala Met His Ser Ser Me
45                150                 155                 160

TGC CAC TGG TTT TCT TTG GAC CAT TTA ATC CTA ACC TAC GCG ACC ATG      588
t Pro Leu Val Phe Phe Gly Pro Phe Asn Pro Asn Leu Arg Asp His As
              165                 170                 175

ACC GGC TGC CCC ATG TCC ATC AGG TAG CCC CCA AGG ACA CAC ATT TGT      636
p Arg Leu Pro His Val His Gln Val Ala Pro Lys Asp Thr His Leu Se
          180                 185                 190

CCC ATG GCA TGG TCT CCT TGA TGT TTC ATT TTA GGT GGA CTT GGA TAG      684
r His Gly Met Val Ser Leu Met Phe His Phe Arg Trp Thr Trp Ile Gl
      195                 200                 205

GAC TGG TCA TCT CAG ATG ATG ATC AGG GTA TTC AGT TTC TCT CAG ATT      732
y Leu Val Ile Ser Asp Asp Asp Gln Gly Ile Gln Phe Leu Ser Asp Le
  210                 215                 220                 2

TAA GAG AAG AAA GCC AAA GGC ATG GGA TCT GTT TGG CTT TTG TTA ATA      780
u Arg Glu Glu Ser Gln Arg His Gly Ile Cys Leu Ala Phe Val Asn Me
25                230                 235                 240

TGA TCC CAG AAA ACA TGC AGA TAT ACA TGA CAA GGG CTA CAA TAT ATG      828
t Ile Pro Glu Asn Met Gln Ile Tyr Met Thr Arg Ala Thr Ile Tyr As
              245                 250                 255

ATA CAC AAA TTA TGA CAT CTT CAG CAA AGG TTG TTA TCA TTT ATG GTG      876
p Thr Gln Ile Met Thr Ser Ser Ala Lys Val Val Ile Ile Tyr Gly As
          260                 265                 270

ACA TGA ACT CTA CTC TAG AAG CAA GCT TTA GAA GAT GGG AAG AGT TAG      924
p Met Asn Ser Thr Leu Glu Ala Ser Phe Arg Arg Trp Glu Glu Leu Gl
      275                 280                 285

GTG CTC GGA GAA TCT GGA TCA CAA CCA CAC AAT GGG ATG TCA TCA CAA      972
y Ala Arg Arg Ile Trp Ile Thr Thr Thr Gln Trp Asp Val Ile Thr As
  290                 295                 300                 3

ATA AAA AAA GAC TTC ACC CT TAATCTCTTC CATGGGACTA TTACTTTTGC ACACC   1027
n Lys Lys Arg Leu His Pro
05                310

ACAAAGATGA GATTCCTAAA TTTAGGAATT TTATGCAAAC AAAGAAAACT GCCAAATACC   1087
TTGTAGATAT TTCTCATACT ATTTTGGAGT GGAATTATTT TAATTGTTCA ATCTCTAAGA   1147
ACAGCAGTAA AATGGGTCAT TTTACATTCA ACAACACATT GCAATGGACA GCACTGCACA   1207
ACTATGATAT GGCCCTGAGC GATGAAGGTT ACAATTTGTA TAATGCTGTT TATGCTGTGG   1267
CCCACACCTA CCATGAATAC ATTCTTCAAC AAGTAGAGTC TCAGAAAAAG GCAAAACCCA   1327
AAAGATATTT CACTGCTTGT CAGCAGGTTC CCTCCTCTGT GTGTAGTGTG GCATGTACTG   1387
CAGGATTCAG GAAAATTCAT CAGAAAGAAA CGGCAGATTG CTGCTTTGAT TGTGTTCAGT   1447
```

```
GCCTAGAAAA TGAGGTTTCC AATGAAACAG ATATGGAACA GTGTGTGAGA TGTCCAGATA   1507
ATAAATATGC CAATTTAGAG CAAACCCACT GCCTCCAAAG AACGGTGTCA TTTCTGGCTT   1567
ATGAAGATCC ATTGGGGATG GCTCTAGGCT GCATGGCACT GTCCTTCTCG GCCATCACAA   1627
TTCTAGTCCT CGTCACATTT GTGAAGTACA AGGATACTCC CATTGTGAAG GCCAATAACC   1687
GCATTCTCAG CTACATCCTG CTCATCTCTC TCGTCTTCTG CTTTCTCTGT TCCCTGCTCT   1747
TCATTGGACA TCCCGACCAG GTCACCTGCA TCTTGCAGCA GACCACATTT GGAGTATTGT   1807
TCACTGTGTC TGTTTCTACA GTGTTGGCCA AAACAATAAC TGTGGTCATG GCTTTCAAGC   1867
TCACTACTCC AGGAAGAAGG ATGAGAGGGA TGATGATGAC AGGGGCACCT AAGTTGGTCA   1927
TTCCCATTTG TACCCTGATC CAACTTGTTC TCTGTGGAAT CTGGTTGGTC ACATCTCCTC   1987
CCTTTATTGA CAGAGATATA CAATCTGAAC ATGGGAAGAT TGTCATTCTT TGCAATAAAG   2047
GCTCTGTCGT TGCCTTCCAC GTCGTCCTGG GATACTTGGG CTCCTTGGCT CTGGGGAGCT   2107
TCACTTTGGC TTTCTTGGCT AGGAACCTTC CTGACACATT CAATGAAGCC AAGTTCCTAA   2167
CTTTCAGCAT GCTGGTGTTC TGCAGTGTCT GGATCACCTT CCTCCCTGTC TACCACAGCA   2227
CCAGGGGGAA GGTCATGGTG GTTGTGGAGG TTTTCTCCAT CTTGGCTTCT AGTGCAGGGT   2287
TGCTAATGTG TATCTTTGTC CCAAAGTGTT ATGTTATTTT AATTAGACCA GATTCAAATT   2347
TTATACAGAA CCACAAAGGT AAATTGCTTT ATTGAAACTT TCATGGTATG AAAATGTTAG   2407
ATGATATTCA ACTTATCTTA TTCTTCATCT TAATAAAAGC AGTACTTCAT CATATAAAAA   2467
ATAAAGTAAT ATACAGATTT ATACTTACAA ACTGGACAGC AAACATGAAT ATGTTGAGAA   2527
CTGGGATTCT CAATTGAGGA ATGGCTACCA ATATTTTGAT CTGTGGTTTT GTGTTTAAGC   2587
CATGTACTTA ATTAATGATT AACATGAGGT TACCCTACTG TCTTTGAACA GCGCCACCTC   2647
TAGGCATGCT GTCCTTGAGT TATAAGAAAG GGTACTGCAT ACACAATGGA CATGAAGCCA   2707
GTAATCAACA TTATTCCACT TGCTTTCATG GAGTTCTTAC TTCCAAGTTC ATGCCTTGAC   2767
TTTATTCAAT GTTCTATGAC AAAGGTAGAA TAAATAAATA AACACTTTCC TCAC         2821
```

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 311 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```
Met Lys Gln Leu Cys Thr Phe Thr Ile Ser Leu Leu Phe Leu Lys Phe
 1               5                  10                  15
Ser Leu Ile Leu Cys Cys Trp Ser Glu Pro Ser Cys Phe Trp Arg Ile
            20                  25                  30
Lys Lys Ser Glu Asp Asn Asp Gly Asp Leu Gln Arg Glu Cys His Phe
        35                  40                  45
Tyr Leu Trp Lys Thr Asp Glu Pro Ile Glu Asp Ser Phe Tyr Asn Tyr
    50                  55                  60
Asp Leu Ser Phe Arg Ile Ala Gly Ser Glu Tyr Glu Leu Leu Leu Val
65                  70                  75                  80
Met Phe Phe Ala Thr Asp Glu Ile Asn Lys Asn Pro Tyr Leu Leu Pro
                85                  90                  95
Asn Met Ser Leu Met Phe Ser Ile Ile Gly Gly Asn Cys His Asp Leu
            100                 105                 110
Leu Arg Ser Leu Asp Gln Glu Tyr Ala Gln Ile Asp Gly His Met Asn
        115                 120                 125
Phe Val Asn Tyr Phe Cys Tyr Leu Asp Asp Ser Cys Ala Thr Gly Leu
    130                 135                 140
Thr Gly Pro Ser Trp Lys Thr Ser Leu Lys Leu Ala Met His Ser Ser
145                 150                 155                 160
Met Pro Leu Val Phe Phe Gly Pro Phe Asn Pro Asn Leu Arg Asp His
                165                 170                 175
Asp Arg Leu Pro His Val His Gln Val Ala Pro Lys Asp Thr His Leu
            180                 185                 190
Ser His Gly Met Val Ser Leu Met Phe His Phe Arg Trp Thr Trp Ile
        195                 200                 205
Gly Leu Val Ile Ser Asp Asp Asp Gln Gly Ile Gln Phe Leu Ser Asp
    210                 215                 220
```

```
Leu Arg Glu Glu Ser Gln Arg His Gly Ile Cys Leu Ala Phe Val Asn
225                 230                 235                 240
Met Ile Pro Glu Asn Met Gln Ile Tyr Met Thr Arg Ala Thr Ile Tyr
                245                 250                 255
Asp Thr Gln Ile Met Thr Ser Ser Ala Lys Val Val Ile Ile Tyr Gly
            260                 265                 270
Asp Met Asn Ser Thr Leu Glu Ala Ser Phe Arg Arg Trp Glu Glu Leu
        275                 280                 285
Gly Ala Arg Arg Ile Trp Ile Thr Thr Thr Gln Trp Asp Val Ile Thr
    290                 295                 300
Asn Lys Lys Arg Leu His Pro
305                 310
```

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2773 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 3...1238
(D) OTHER INFORMATION: VR13

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```
AA GCA AGT TGC TTT TGG CGG ATA AAG AAT AGT GAA GAT AAT GAT GGA      47
   Ala Ser Cys Phe Trp Arg Ile Lys Asn Ser Glu Asp Asn Asp Gly
    1               5                  10                  15

GAT TTG CAA AGG GAA TGT CAT TTT TAC CTT GGG GCA GTT GAT AAA CCA      95
Asp Leu Gln Arg Glu Cys His Phe Tyr Leu Gly Ala Val Asp Lys Pro
                20                  25                  30

ATT GAA GAT AAT TTT TAT AAT TCA CTT TTA AAG TTT AGA ATT GCA GCA     143
Ile Glu Asp Asn Phe Tyr Asn Ser Leu Leu Lys Phe Arg Ile Ala Ala
            35                  40                  45

AGT GAA TAT GAG TTT CTT CTG GTA ATG TTT TTT GCT ACT GAT GAG ATC     191
Ser Glu Tyr Glu Phe Leu Leu Val Met Phe Phe Ala Thr Asp Glu Ile
        50                  55                  60

AAC AAG AAT CCT TAT CTT TTA CCC AAC ATA ACT TTG ATG TTC TCC ATC     239
Asn Lys Asn Pro Tyr Leu Leu Pro Asn Ile Thr Leu Met Phe Ser Ile
    65                  70                  75

ATT GGT GGA AAC TGT CAT GAT TTA TTG AGA GGT TTG GAT CAA GCA TAT     287
Ile Gly Gly Asn Cys His Asp Leu Leu Arg Gly Leu Asp Gln Ala Tyr
80                  85                  90                  95

ACA CAA ATA AAT GGA CAT ATG AAT TTT GTT AAT TAT TTC TGT TAT TTA     335
Thr Gln Ile Asn Gly His Met Asn Phe Val Asn Tyr Phe Cys Tyr Leu
                100                 105                 110

GAT GAT TCA TGT GCC ATA GGT CTT ACA GGA CCA TCA TGG AAA ACA TCC     383
Asp Asp Ser Cys Ala Ile Gly Leu Thr Gly Pro Ser Trp Lys Thr Ser
            115                 120                 125

TTA AAA CTG GCA ATG CAT TCT TCA ATG CCA CTG GTT TTC TTT GGA TCA     431
Leu Lys Leu Ala Met His Ser Ser Met Pro Leu Val Phe Phe Gly Ser
```

```
        130                 135                 140

TTT AAT CCT AAC CTA CAT GAC CAT GAC CGG CTG CAC CAT GTC CAT CAA      479
Phe Asn Pro Asn Leu His Asp His Asp Arg Leu His His Val His Gln
    145                 150                 155

GTA GCC ACC AAG GAC ACA CAT TTG TCC CAT GGC ATT GTC TCC TTG ATG      527
Val Ala Thr Lys Asp Thr His Leu Ser His Gly Ile Val Ser Leu Met
160                 165                 170                 175

TTT CAT TTT AGA TGG ACT TGG ATA GGA CTG GTC ATC TCA GAT GAT GAC      575
Phe His Phe Arg Trp Thr Trp Ile Gly Leu Val Ile Ser Asp Asp Asp
                180                 185                 190

AAG GGT ATT CAG TTT CTC TCA GAT TTA AGA GAA GAA AGC CAA AGG CAT      623
Lys Gly Ile Gln Phe Leu Ser Asp Leu Arg Glu Glu Ser Gln Arg His
            195                 200                 205

GGG ATC TGT TTA GCT TTT GTT AAT ATG ATC CCA GAA AAC ATG CAG ATA      671
Gly Ile Cys Leu Ala Phe Val Asn Met Ile Pro Glu Asn Met Gln Ile
        210                 215                 220

TAC ATG ACA AGG GCT ACA ATA TAT GAT AAA CAA ATT ATG ACG TCT TTA      719
Tyr Met Thr Arg Ala Thr Ile Tyr Asp Lys Gln Ile Met Thr Ser Leu
    225                 230                 235

GCA AAA GTT GTT ATC ATT TAT GGT GAA ATG AAC TCT ACA CTA GAA GTA      767
Ala Lys Val Val Ile Ile Tyr Gly Glu Met Asn Ser Thr Leu Glu Val
240                 245                 250                 255

AGC TTT AGA AGA TGG GAA AAT TTA GGT GCT CGG AGA ATC TGG ATC ACA      815
Ser Phe Arg Arg Trp Glu Asn Leu Gly Ala Arg Arg Ile Trp Ile Thr
                260                 265                 270

ACC TCA CAA TGG GAT GTC ATC ACA AAT AAA AAA GAA TTC ACC CTT AAT      863
Thr Ser Gln Trp Asp Val Ile Thr Asn Lys Lys Glu Phe Thr Leu Asn
            275                 280                 285

CTC TTC CAT GGG ACT ATT ACT TTT GCA CAC CGC AGA TTT GAG ATT CCT      911
Leu Phe His Gly Thr Ile Thr Phe Ala His Arg Arg Phe Glu Ile Pro
        290                 295                 300

AAA TTT AAA AAA TTT ATG CAA ACA ATG AAC ACT GCC AAA TAC CCA GTA      959
Lys Phe Lys Lys Phe Met Gln Thr Met Asn Thr Ala Lys Tyr Pro Val
    305                 310                 315

GAT ATT TCT CAT ACT ATA TTG GAG TGG AAT TAT TTT AAT TGT TCA ATC     1007
Asp Ile Ser His Thr Ile Leu Glu Trp Asn Tyr Phe Asn Cys Ser Ile
320                 325                 330                 335

TCT AAG AAC AGC AGT AAA ATG GAT CAT ATT ACA TTC AAC AAC ACA TTG     1055
Ser Lys Asn Ser Ser Lys Met Asp His Ile Thr Phe Asn Asn Thr Leu
                340                 345                 350

GAA TGG ACA GCA CTG CAC AAC TAT GAT ATG GTG ATG AGT GAT GAA GGT     1103
Glu Trp Thr Ala Leu His Asn Tyr Asp Met Val Met Ser Asp Glu Gly
            355                 360                 365

TAC AAT TTG TAT AAT GCT GTT TAT GCT GTG GCC CAC ACC TAC CAT GAA     1151
Tyr Asn Leu Tyr Asn Ala Val Tyr Ala Val Ala His Thr Tyr His Glu
        370                 375                 380

CAT ATT TTT CAA CAA GTA GAG TCT CAG AAA AAG GCA AAA CCC AAA AGA     1199
His Ile Phe Gln Gln Val Glu Ser Gln Lys Lys Ala Lys Pro Lys Arg
    385                 390                 395
```

```
TTT TTC ACT GTT TGT CAG CAG CAG ATA TGG AAC AGT GTG TGAAGTGTCC AT  1250
Phe Phe Thr Val Cys Gln Gln Gln Ile Trp Asn Ser Val
400                 405                 410

ATGATAAGTA TGCCAACATA GAGAAAACCC ACTGCCTCTC AAGAGCTGTA TCATTTCTGG   1310
CTTATGAAGA TCCATTGGGG ATAGCTCTAG GCTGCATAGC ACTGTCCTTC TCAGCCATCA   1370
CAATTCTAGT ACTAATCACA TTTTTGAAGT ACAAGGATAC TCCCATTGTG AAGGCCAATA   1430
ACCGCATTCT CAGCTACATC CTGCTCATCT CTCTAGTCTT CTGCTTTCTC TGCTCCCTGC   1490
TCTTCATTGG ACATCCAAAC CAGGTCTCCT GCGTCTTGCA GCAGACCACA TTTGGAGTAT   1550
TTTTCACTGT GTCTGTTTCT ACAGTGTTGG CCAAAACAAT AACTGTGGTC ATGGCTTTCA   1610
AGCTCACTAC TCCAGGAAGA AGAATGAGAG AGATGTTGGT AACAGGGGCA CCTAAGTTGG   1670
TCATTCCCAT TTGTACCCTA ATCCAATTTG TTCTCTGTGG AATCTGGTTG ATAACATCTC   1730
CTCCATTTAT TGACAGAGAT ATACAATCTG AGCATGGGAA GATTGTCATT CTTTGCAATA   1790
AAGGCTCTGT CATTGCCTTC CATGTTGTCC TGGGATACTT GGGCTCCTTG GCTCTGGGGA   1850
GCTTCACTTT GGCTTTCTTG GCTAGGAACC TTCCTGACAC ATTCAATGAA GCCAAATTCC   1910
TGACTTTCAG CATGCTGGTG TTCTGCAGTG TCTGGATCAC CTTTCTCCCT GTCTACCATA   1970
GCACCAGGGG GAAGGTCATG GTGGTTGTGG AGGTTTTCTC AATCTTGGCT TCTAGTGCAG   2030
GGTTGCTAAT GTGTATCTTT GTCCCAAAGT GTTATGTTAT TTTAGTTAGA CCAGATTCAA   2090
ATTTTATACG GAAGTACAAA GATAAATTTC GTTATTGAAA TATTCATACT ATGAAAATGT   2150
TAGATTATAC TCAACATATT TTTCTTTGTC TTAACAAAAG TAGTACTTAA TCTTATAAAA   2210
ATTTAAATAA TATACAAATT TGAACTTACA AACAGGACAG AACTGTCTAT TGTAATACCA   2270
ATTACAAAAC TTTGGTGAAA AATGGTCTCA TTCATAAGGA CACAATTCTG AAGATATTGA   2330
GAACCAGGAA TCTCAACTGC GGAAACGCTA CCATCATCCT GACCTGTGGT TTTGTGTGTA   2390
AAGCATGAAC TTAATTAATG ATTAATATAA GGTGACCATA CTGACTGTGA ACACTACCAT   2450
CTCTGGGCAA GTTGTTCTTG TAGTTGTAAG AAAAAGCTCT GAAGACAACA TGGAAGTAAA   2510
GCCAGTAATC ACCATTATCC CTCATGCTTT CATGGAGTGG CTGCATCCAA TTTCATGCCT   2570
TGGCTTCATT CAATATACTG TGACCAAGGT ACATAAGTAA AGAAACACTT TTCTTACAAG   2630
CTTCTTCTGA TCGTTGTGGG TTTTTTTGTT TTTTGTTTTT TGTTTTTTGT TTGTTTGTTT   2690
GTATTTTTAC ATCAACGGAA TTTAAAATAT CAACAAAATG GTAAATTGTT TCTGTTGAGA   2750
TTTAGAATAT CATCGATTCC TGA                                           2773
```

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 412 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```
Ala Ser Cys Phe Trp Arg Ile Lys Asn Ser Glu Asp Asn Asp Gly Asp
 1               5                  10                  15
Leu Gln Arg Glu Cys His Phe Tyr Leu Gly Ala Val Asp Lys Pro Ile
            20                  25                  30
Glu Asp Asn Phe Tyr Asn Ser Leu Leu Lys Phe Arg Ile Ala Ala Ser
        35                  40                  45
Glu Tyr Glu Phe Leu Leu Val Met Phe Phe Ala Thr Asp Glu Ile Asn
    50                  55                  60
Lys Asn Pro Tyr Leu Leu Pro Asn Ile Thr Leu Met Phe Ser Ile Ile
65                  70                  75                  80
Gly Gly Asn Cys His Asp Leu Leu Arg Gly Leu Asp Gln Ala Tyr Thr
                85                  90                  95
Gln Ile Asn Gly His Met Asn Phe Val Asn Tyr Phe Cys Tyr Leu Asp
            100                 105                 110
Asp Ser Cys Ala Ile Gly Leu Thr Gly Pro Ser Trp Lys Thr Ser Leu
        115                 120                 125
Lys Leu Ala Met His Ser Ser Met Pro Leu Val Phe Phe Gly Ser Phe
    130                 135                 140
Asn Pro Asn Leu His Asp His Asp Arg Leu His His Val His Gln Val
145                 150                 155                 160
Ala Thr Lys Asp Thr His Leu Ser His Gly Ile Val Ser Leu Met Phe
```

```
                165                 170                 175
His Phe Arg Trp Thr Trp Ile Gly Leu Val Ile Ser Asp Asp Asp Lys
            180                 185                 190
Gly Ile Gln Phe Leu Ser Asp Leu Arg Glu Glu Ser Gln Arg His Gly
        195                 200                 205
Ile Cys Leu Ala Phe Val Asn Met Ile Pro Glu Asn Met Gln Ile Tyr
    210                 215                 220
Met Thr Arg Ala Thr Ile Tyr Asp Lys Gln Ile Met Thr Ser Leu Ala
225                 230                 235                 240
Lys Val Val Ile Ile Tyr Gly Glu Met Asn Ser Thr Leu Glu Val Ser
                245                 250                 255
Phe Arg Arg Trp Glu Asn Leu Gly Ala Arg Arg Ile Trp Ile Thr Thr
            260                 265                 270
Ser Gln Trp Asp Val Ile Thr Asn Lys Lys Glu Phe Thr Leu Asn Leu
        275                 280                 285
Phe His Gly Thr Ile Thr Phe Ala His Arg Arg Phe Glu Ile Pro Lys
    290                 295                 300
Phe Lys Lys Phe Met Gln Thr Met Asn Thr Ala Lys Tyr Pro Val Asp
305                 310                 315                 320
Ile Ser His Thr Ile Leu Glu Trp Asn Tyr Phe Asn Cys Ser Ile Ser
                325                 330                 335
Lys Asn Ser Ser Lys Met Asp His Ile Thr Phe Asn Asn Thr Leu Glu
            340                 345                 350
Trp Thr Ala Leu His Asn Tyr Asp Met Val Met Ser Asp Glu Gly Tyr
        355                 360                 365
Asn Leu Tyr Asn Ala Val Tyr Ala Val Ala His Thr Tyr His Glu His
    370                 375                 380
Ile Phe Gln Gln Val Glu Ser Gln Lys Lys Ala Lys Pro Lys Arg Phe
385                 390                 395                 400
Phe Thr Val Cys Gln Gln Gln Ile Trp Asn Ser Val
                405                 410
```

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3108 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 116...2527
(D) OTHER INFORMATION: VR14

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

```
GAATATGCAA TAAACATCTC CTTTGCCTAA AGAAATAAAA GCTGGTAGAA ATCTGATGTG     60
CTGATATGCA TGGCACTTCA CAATCCACAC TGCCCAGGTT TAAGGCAGGA AAAAG ATG    118
                                                              Met
                                                               1

TTC ATT TTC ATG GAA GTC TTC TTC CTC CTT AAT ATT ACA CTT CTC ATG      166
Phe Ile Phe Met Glu Val Phe Phe Leu Leu Asn Ile Thr Leu Leu Met
          5                  10                  15

GCC AAT TTC ATT GAT CCC AGG TGC TTT TGG AGA ATA AAT TTG GAT GAA      214
Ala Asn Phe Ile Asp Pro Arg Cys Phe Trp Arg Ile Asn Leu Asp Glu
      20                  25                  30

ATA ATG GAT GAA TAT TTG GGA TTA TCT TGT GCT TTC ATC CTG GCA GCA      262
Ile Met Asp Glu Tyr Leu Gly Leu Ser Cys Ala Phe Ile Leu Ala Ala
```

```
    35                  40                  45

GTT CAG ACA CCC ATT GAA AAT GAT TAT TTC AAC AAG ACT CTT AAT GTT      310
Val Gln Thr Pro Ile Glu Asn Asp Tyr Phe Asn Lys Thr Leu Asn Val
50                  55                  60                  65

CTA AAA ACA ACT AAA AAC CAC AAA TAT GCT TTG GCA TTG GTG TTT GCA      358
Leu Lys Thr Thr Lys Asn His Lys Tyr Ala Leu Ala Leu Val Phe Ala
                70                  75                  80

ATG GAT GAA ATC AAC AGA AAT CCT GAT CTT TTA CCA AAT ATG TCT TTG      406
Met Asp Glu Ile Asn Arg Asn Pro Asp Leu Leu Pro Asn Met Ser Leu
            85                  90                  95

ATT ATA AGA TAC ACT TTG GGC CGT TGT GAT GGA AAA ACT GTA ATA CCT      454
Ile Ile Arg Tyr Thr Leu Gly Arg Cys Asp Gly Lys Thr Val Ile Pro
        100                 105                 110

ACA CCA TAT TTA TTT CGT AAA AAA AAA GAA AGC CCT ATC CCT AAT TAT      502
Thr Pro Tyr Leu Phe Arg Lys Lys Lys Glu Ser Pro Ile Pro Asn Tyr
    115                 120                 125

TTC TGT AAT GAA GAG ACT ATG TGT TCC TAT CTG CTT ACA GGA CCC CAT      550
Phe Cys Asn Glu Glu Thr Met Cys Ser Tyr Leu Leu Thr Gly Pro His
130                 135                 140                 145

TGG GAG GTA TCT TTA GGT TTC TGG AAG CAC ATG AAC AGC TTC TTA TCT      598
Trp Glu Val Ser Leu Gly Phe Trp Lys His Met Asn Ser Phe Leu Ser
                150                 155                 160

CCA CGT ATC CTT CAG CTT ACC TAT GGA CCT TTC CAC TCC ATC TTC AGT      646
Pro Arg Ile Leu Gln Leu Thr Tyr Gly Pro Phe His Ser Ile Phe Ser
            165                 170                 175

GAT GAT GAA CAA TAT CCC TAT CTC TAT CAG ATG GCC CCA AAG GAC ACA      694
Asp Asp Glu Gln Tyr Pro Tyr Leu Tyr Gln Met Ala Pro Lys Asp Thr
        180                 185                 190

TCT CTA GCA TTG GCA ATG GTC TCC TTC ATA CTT TAC TTT AGC TGG AAC      742
Ser Leu Ala Leu Ala Met Val Ser Phe Ile Leu Tyr Phe Ser Trp Asn
    195                 200                 205

TGG ATT GGC CTT GTC ATT CCA GAT GAT GAC CAA GGA AAC CAA TTT CTT      790
Trp Ile Gly Leu Val Ile Pro Asp Asp Asp Gln Gly Asn Gln Phe Leu
210                 215                 220                 225

TTA GAG TTG AAG AAA CAG AGT GAA AAC AAG GAA ATT TGC TTT GCC TTT      838
Leu Glu Leu Lys Lys Gln Ser Glu Asn Lys Glu Ile Cys Phe Ala Phe
                230                 235                 240

GTG AAA ATG ATC TCT GTT GAT GAT GTT TCA TTT CCA CAA AAT ACT GAA      886
Val Lys Met Ile Ser Val Asp Asp Val Ser Phe Pro Gln Asn Thr Glu
            245                 250                 255

ATG TAC TAC AAC CAA ATT GTG ATG TCA TCC ACA AAT GTT ATT ATC ATT      934
Met Tyr Tyr Asn Gln Ile Val Met Ser Ser Thr Asn Val Ile Ile Ile
        260                 265                 270

TAT GGA GAA ACA TAC AAT TTC ATT GAT TTG ATC TTC AGA ATG TGG GAA      982
Tyr Gly Glu Thr Tyr Asn Phe Ile Asp Leu Ile Phe Arg Met Trp Glu
    275                 280                 285

CCT CCC ATT TTA CAG AGA ATA TGG ATC ACC ACA AAA CAA TTG AAT TTC     1030
Pro Pro Ile Leu Gln Arg Ile Trp Ile Thr Thr Lys Gln Leu Asn Phe
290                 295                 300                 305
```

```
CCT ACC AGG AAA AAA GAC ATA AGT CAT GGC ACA TTC TAT GGA TCA CTT     1078
Pro Thr Arg Lys Lys Asp Ile Ser His Gly Thr Phe Tyr Gly Ser Leu
                310                 315                 320

ACT TTT CTA CCC CAC CAT GGT GTG ATT TCT GGT TTT AAA AAT TTT GTA     1126
Thr Phe Leu Pro His His Gly Val Ile Ser Gly Phe Lys Asn Phe Val
            325                 330                 335

CAG ACA TGG TTC CAT CTC AGA AAC ACA GAT TTA TAT CTA GTA ATG CAA     1174
Gln Thr Trp Phe His Leu Arg Asn Thr Asp Leu Tyr Leu Val Met Gln
        340                 345                 350

GAG TGG AAA TAC TTT AAC TAT GAA GAC TCA GCA TCT ACC TGT AAA ATA     1222
Glu Trp Lys Tyr Phe Asn Tyr Glu Asp Ser Ala Ser Thr Cys Lys Ile
    355                 360                 365

CTG AAG AAC AAT TCA TCT AAT GCC TCA TTT GAT TGG CTA ATG GAA CAG     1270
Leu Lys Asn Asn Ser Ser Asn Ala Ser Phe Asp Trp Leu Met Glu Gln
370                 375                 380                 385

AAG TTT GAC ATG ACC TTT AGT GAG AAT AGT CAT AAC ATA TAC AAT GCT     1318
Lys Phe Asp Met Thr Phe Ser Glu Asn Ser His Asn Ile Tyr Asn Ala
                390                 395                 400

GTG CAT GCC ATA GCC CAT GCC CTC CAT GAG ATG AAT CTG CAA CAG GCT     1366
Val His Ala Ile Ala His Ala Leu His Glu Met Asn Leu Gln Gln Ala
            405                 410                 415

GAT AAT CAG GCA ATA GAC AAT GGG AAA AAG GAG CCC AGT TCC TCC CAC     1414
Asp Asn Gln Ala Ile Asp Asn Gly Lys Lys Glu Pro Ser Ser Ser His
        420                 425                 430

TGC TTG AAG GTA AAC TCC TTT CTA AGA AGG ATT TAC TTC ACT AAT CCT     1462
Cys Leu Lys Val Asn Ser Phe Leu Arg Arg Ile Tyr Phe Thr Asn Pro
    435                 440                 445

CCT GGG GAC AAA GTG TTT ATG AAG CAA AGA GTA ATA ATG CAC GAT GAA     1510
Pro Gly Asp Lys Val Phe Met Lys Gln Arg Val Ile Met His Asp Glu
450                 455                 460                 465

TAT GAC ATT GTT CAC TTT GTG AAT CTC TCA CAA CAC CTT GGG ATT AAG     1558
Tyr Asp Ile Val His Phe Val Asn Leu Ser Gln His Leu Gly Ile Lys
                470                 475                 480

ATG AAG TTA GGA AAG TTC AGC CCA TAT TTA CCA CAT GGT CGA CAC TCT     1606
Met Lys Leu Gly Lys Phe Ser Pro Tyr Leu Pro His Gly Arg His Ser
            485                 490                 495

CAC TTA TAT GTA GAC AGG ATT GAG TTG GCC ACA GGA AGA AGA AAG ATG     1654
His Leu Tyr Val Asp Arg Ile Glu Leu Ala Thr Gly Arg Arg Lys Met
        500                 505                 510

CCA TCC TCT GTG TGC AGT GCT GAT TGT AGT CCT GGA TTC AGA AGA TTA     1702
Pro Ser Ser Val Cys Ser Ala Asp Cys Ser Pro Gly Phe Arg Arg Leu
    515                 520                 525

TGG AAG GAG GGA ATG GCA GCC TGC TGT TTT GTT TGC AGC CCC TGC CCT     1750
Trp Lys Glu Gly Met Ala Ala Cys Cys Phe Val Cys Ser Pro Cys Pro
530                 535                 540                 545

GAA AAT GAA ATT TCT AAT GAG ACA ACT GTG GTA CTT TGT GTC TTT GTG     1798
Glu Asn Glu Ile Ser Asn Glu Thr Thr Val Val Leu Cys Val Phe Val
                550                 555                 560

AAG CAT CAT GAC ACT CCT ATT GTG AAG GCC AAT AAC AGA AGC CTC AGC     1846
```

```
Lys His His Asp Thr Pro Ile Val Lys Ala Asn Asn Arg Ser Leu Ser
            565                 570                 575

TAC CTA TTA CTC ATG TCA CTC ATG TCC TGT TTT CTG TGC TCC TTT TTC    1894
Tyr Leu Leu Leu Met Ser Leu Met Ser Cys Phe Leu Cys Ser Phe Phe
        580                 585                 590

TTC ATT GGC CTT CCA AAC AGA GCC ATC TGT GTC TTA CAG CAA ATC ACA    1942
Phe Ile Gly Leu Pro Asn Arg Ala Ile Cys Val Leu Gln Gln Ile Thr
    595                 600                 605

TTT GGA ATT GTA TTC ACT ATG GCT GTT TCC ACA GTT CTG GCC AAA ACA    1990
Phe Gly Ile Val Phe Thr Met Ala Val Ser Thr Val Leu Ala Lys Thr
610                 615                 620                 625

GTC ACT GTG GTT CTG GCT TTC AAA GTC ACA GAC CCA GGA AGA AGA TTG    2038
Val Thr Val Val Leu Ala Phe Lys Val Thr Asp Pro Gly Arg Arg Leu
                630                 635                 640

AGA AAC TTC CTG GTA TCA GGA ACA CCC AAC TAC ATT ATT CCC ATA TGT    2086
Arg Asn Phe Leu Val Ser Gly Thr Pro Asn Tyr Ile Ile Pro Ile Cys
            645                 650                 655

TCC CTA CTC CAA TGT GTT CTG TGT GCA ATC TGG CTA GCA GTT TCT CCT    2134
Ser Leu Leu Gln Cys Val Leu Cys Ala Ile Trp Leu Ala Val Ser Pro
        660                 665                 670

CCC TTT GTT GAT ATT GAT GAA CAC ACT CTC CAT GGC CAC ATC ATC ATT    2182
Pro Phe Val Asp Ile Asp Glu His Thr Leu His Gly His Ile Ile Ile
    675                 680                 685

GTG TGC AAC AAG GGC TCA GTT ACT GCA TTC TAC TGT ATC CTA GGA TAC    2230
Val Cys Asn Lys Gly Ser Val Thr Ala Phe Tyr Cys Ile Leu Gly Tyr
690                 695                 700                 705

TTG GCC TGC CTG GCA CTT GGA AAC TTC TCT GTG GCT TTC TTG GCC AAG    2278
Leu Ala Cys Leu Ala Leu Gly Asn Phe Ser Val Ala Phe Leu Ala Lys
                710                 715                 720

AAT CTG CCT GAC ACA TTC AAT GAA GCC AAG TTC TTG ACC TTC AGC ATG    2326
Asn Leu Pro Asp Thr Phe Asn Glu Ala Lys Phe Leu Thr Phe Ser Met
            725                 730                 735

CTA GTG TTC TGT AGT GTC TGG GTC ACC TTC CTC CCT GTC TAC CAT AGC    2374
Leu Val Phe Cys Ser Val Trp Val Thr Phe Leu Pro Val Tyr His Ser
        740                 745                 750

ACC AAG GGC AAA CAC ATG GTT GCT GTG GAG ATC TTC TCC ATC TTG GCA    2422
Thr Lys Gly Lys His Met Val Ala Val Glu Ile Phe Ser Ile Leu Ala
    755                 760                 765

TCC AGT GCT GGG ATC CTT GGA TGT ATA TTT GTA CCC AAG ATT TAT ATC    2470
Ser Ser Ala Gly Ile Leu Gly Cys Ile Phe Val Pro Lys Ile Tyr Ile
770                 775                 780                 785

ATT TTA ATG AGA CCA GAG AGA AAT TCG ACC CAA AAG ATC AGG GAA AAA    2518
Ile Leu Met Arg Pro Glu Arg Asn Ser Thr Gln Lys Ile Arg Glu Lys
                790                 795                 800

TCA TAT TTC TGAACAAATA TTTAGGAATT CTGTCAAATG TAAAGTTGGT ACATACCCA  2576
Ser Tyr Phe

CCAAATATTG GGTTATAGTG CATGTGTCTA GTTTTAGAAT CACTCTCACT GGTTGCTCTA  2636
GTGATATCAG CAAGTATCAT ATCTACTGAA CTTCCCTACA GTGTCCATAA AATCTTGTAC  2696
```

```
TCATTCACTT TCTTCATTTT CTCTCAGAGA ACTAAACTCT CTAATTATTA CAATTTTATT   2756
CTTCATTTTG CTTTCATGGA GATTGCCCTC TGGTAACTTC CAAAAAATGT TGATAAGGCA   2816
GTTGAATCCA CCACTTTGTG TAGAAAAATG AGATCTAGGA AGACAGGGTT ACACATAAAA   2876
ACCATCTACC AAAATAAATA ATCAATGAGA AACACAGACT AACTAAATAA TCAGCAAAGA   2936
TGAAATCAGA ACATATTTTC TAATTTCCAG TAAGAGCACA CACATAAGAA AATACTTACT   2996
TTTTTCATCT GTTCTTCAAT CTACTGGCCA ATAGTCTAAG GAGGAAATGT TCCTTTTCTG   3056
CTGTCAAATA AAAATATATT ATATCCAAAA AAAAAAAAAA AAAAAAAAAA AA            3108
```

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 804 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

```
Met Phe Ile Phe Met Glu Val Phe Phe Leu Leu Asn Ile Thr Leu Leu
 1               5                  10                  15
Met Ala Asn Phe Ile Asp Pro Arg Cys Phe Trp Arg Ile Asn Leu Asp
            20                  25                  30
Glu Ile Met Asp Glu Tyr Leu Gly Leu Ser Cys Ala Phe Ile Leu Ala
        35                  40                  45
Ala Val Gln Thr Pro Ile Glu Asn Asp Tyr Phe Asn Lys Thr Leu Asn
    50                  55                  60
Val Leu Lys Thr Thr Lys Asn His Lys Tyr Ala Leu Ala Leu Val Phe
65                  70                  75                  80
Ala Met Asp Glu Ile Asn Arg Asn Pro Asp Leu Leu Pro Asn Met Ser
                85                  90                  95
Leu Ile Ile Arg Tyr Thr Leu Gly Arg Cys Asp Gly Lys Thr Val Ile
            100                 105                 110
Pro Thr Pro Tyr Leu Phe Arg Lys Lys Lys Glu Ser Pro Ile Pro Asn
        115                 120                 125
Tyr Phe Cys Asn Glu Glu Thr Met Cys Ser Tyr Leu Leu Thr Gly Pro
    130                 135                 140
His Trp Glu Val Ser Leu Gly Phe Trp Lys His Met Asn Ser Phe Leu
145                 150                 155                 160
Ser Pro Arg Ile Leu Gln Leu Thr Tyr Gly Pro Phe His Ser Ile Phe
                165                 170                 175
Ser Asp Asp Glu Gln Tyr Pro Tyr Leu Tyr Gln Met Ala Pro Lys Asp
            180                 185                 190
Thr Ser Leu Ala Leu Ala Met Val Ser Phe Ile Leu Tyr Phe Ser Trp
        195                 200                 205
Asn Trp Ile Gly Leu Val Ile Pro Asp Asp Asp Gln Gly Asn Gln Phe
    210                 215                 220
Leu Leu Glu Leu Lys Lys Gln Ser Glu Asn Lys Glu Ile Cys Phe Ala
225                 230                 235                 240
Phe Val Lys Met Ile Ser Val Asp Asp Val Ser Phe Pro Gln Asn Thr
                245                 250                 255
Glu Met Tyr Tyr Asn Gln Ile Val Met Ser Ser Thr Asn Val Ile Ile
            260                 265                 270
Ile Tyr Gly Glu Thr Tyr Asn Phe Ile Asp Leu Ile Phe Arg Met Trp
        275                 280                 285
Glu Pro Pro Ile Leu Gln Arg Ile Trp Ile Thr Thr Lys Gln Leu Asn
    290                 295                 300
Phe Pro Thr Arg Lys Lys Asp Ile Ser His Gly Thr Phe Tyr Gly Ser
305                 310                 315                 320
Leu Thr Phe Leu Pro His His Gly Val Ile Ser Gly Phe Lys Asn Phe
                325                 330                 335
Val Gln Thr Trp Phe His Leu Arg Asn Thr Asp Leu Tyr Leu Val Met
            340                 345                 350
```

```
Gln Glu Trp Lys Tyr Phe Asn Tyr Glu Asp Ser Ala Ser Thr Cys Lys
        355                 360                 365
Ile Leu Lys Asn Asn Ser Ser Asn Ala Ser Phe Asp Trp Leu Met Glu
    370                 375                 380
Gln Lys Phe Asp Met Thr Phe Ser Glu Asn Ser His Asn Ile Tyr Asn
385                 390                 395                 400
Ala Val His Ala Ile Ala His Ala Leu His Glu Met Asn Leu Gln Gln
                405                 410                 415
Ala Asp Asn Gln Ala Ile Asp Asn Gly Lys Lys Glu Pro Ser Ser Ser
            420                 425                 430
His Cys Leu Lys Val Asn Ser Phe Leu Arg Arg Ile Tyr Phe Thr Asn
        435                 440                 445
Pro Pro Gly Asp Lys Val Phe Met Lys Gln Arg Val Ile Met His Asp
    450                 455                 460
Glu Tyr Asp Ile Val His Phe Val Asn Leu Ser Gln His Leu Gly Ile
465                 470                 475                 480
Lys Met Lys Leu Gly Lys Phe Ser Pro Tyr Leu Pro His Gly Arg His
                485                 490                 495
Ser His Leu Tyr Val Asp Arg Ile Glu Leu Ala Thr Gly Arg Arg Lys
            500                 505                 510
Met Pro Ser Ser Val Cys Ser Ala Asp Cys Ser Pro Gly Phe Arg Arg
        515                 520                 525
Leu Trp Lys Glu Gly Met Ala Ala Cys Cys Phe Val Cys Ser Pro Cys
    530                 535                 540
Pro Glu Asn Glu Ile Ser Asn Glu Thr Thr Val Val Leu Cys Val Phe
545                 550                 555                 560
Val Lys His His Asp Thr Pro Ile Val Lys Ala Asn Asn Arg Ser Leu
                565                 570                 575
Ser Tyr Leu Leu Leu Met Ser Leu Met Ser Cys Phe Leu Cys Ser Phe
            580                 585                 590
Phe Phe Ile Gly Leu Pro Asn Arg Ala Ile Cys Val Leu Gln Gln Ile
        595                 600                 605
Thr Phe Gly Ile Val Phe Thr Met Ala Val Ser Thr Val Leu Ala Lys
    610                 615                 620
Thr Val Thr Val Val Leu Ala Phe Lys Val Thr Asp Pro Gly Arg Arg
625                 630                 635                 640
Leu Arg Asn Phe Leu Val Ser Gly Thr Pro Asn Tyr Ile Ile Pro Ile
                645                 650                 655
Cys Ser Leu Leu Gln Cys Val Leu Cys Ala Ile Trp Leu Ala Val Ser
            660                 665                 670
Pro Pro Phe Val Asp Ile Asp Glu His Thr Leu His Gly His Ile Ile
        675                 680                 685
Ile Val Cys Asn Lys Gly Ser Val Thr Ala Phe Tyr Cys Ile Leu Gly
    690                 695                 700
Tyr Leu Ala Cys Leu Ala Leu Gly Asn Phe Ser Val Ala Phe Leu Ala
705                 710                 715                 720
Lys Asn Leu Pro Asp Thr Phe Asn Glu Ala Lys Phe Leu Thr Phe Ser
                725                 730                 735
Met Leu Val Phe Cys Ser Val Trp Val Thr Phe Leu Pro Val Tyr His
            740                 745                 750
Ser Thr Lys Gly Lys His Met Val Ala Val Glu Ile Phe Ser Ile Leu
        755                 760                 765
Ala Ser Ser Ala Gly Ile Leu Gly Cys Ile Phe Val Pro Lys Ile Tyr
    770                 775                 780
Ile Ile Leu Met Arg Pro Glu Arg Asn Ser Thr Gln Lys Ile Arg Glu
785                 790                 795                 800
Lys Ser Tyr Phe
```

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3689 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 39...419
(D) OTHER INFORMATION: VR15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```
TCAAAATCCG CACTGCCCAA GTTTAAGGCA GGAAAAAT ATG TTC ATT TTC ATG GGA    56
                                          Met Phe Ile Phe Met Gly
                                           1               5

GTC TTC TTC CTC CTT AAT ATT ACA CTT CTC ATG GCC AAT TTC ATT AAT     104
Val Phe Phe Leu Leu Asn Ile Thr Leu Leu Met Ala Asn Phe Ile Asn
            10                  15                  20

CCC AGG TGC TTT TGG AGA ATA AAT TTG GAT GAA ATA ACG GAT GAA TAT     152
Pro Arg Cys Phe Trp Arg Ile Asn Leu Asp Glu Ile Thr Asp Glu Tyr
        25                  30                  35

TTG GGA TTA TCT TGT ACT TTC ATC CTG GCG GCA GTT CAG ACA CCC ACT     200
Leu Gly Leu Ser Cys Thr Phe Ile Leu Ala Ala Val Gln Thr Pro Thr
    40                  45                  50

GAA AAA GAT TAT TTC AAC AAG ACT CTT AAT GTT CTA AAA ACA ACT AAA     248
Glu Lys Asp Tyr Phe Asn Lys Thr Leu Asn Val Leu Lys Thr Thr Lys
55                  60                  65                  70

AAC CAC AAA TAT GCT TTG GCA TTG GTG TTT GCA ATG GAT GAA ATC AAC     296
Asn His Lys Tyr Ala Leu Ala Leu Val Phe Ala Met Asp Glu Ile Asn
                75                  80                  85

AGA AAT CCT GAT CTT TTA CCA AAT ATG TCT TTG ATT ATA AGA TAC ACT     344
Arg Asn Pro Asp Leu Leu Pro Asn Met Ser Leu Ile Ile Arg Tyr Thr
            90                  95                  100

TTG GGC CTT TGT GAT GGA AAA ACT GTA ACA CCT ACA CCA TAT TTA TTT     392
Leu Gly Leu Cys Asp Gly Lys Thr Val Thr Pro Thr Pro Tyr Leu Phe
        105                 110                 115

CAT AAA AAA AAA ACA AAG CCC TAT CCC TAATTATTTC TGTAATGAAG AGACTAT   446
His Lys Lys Lys Thr Lys Pro Tyr Pro
    120                 125

GTGTTCATTT CTGCTTTCAG GACCCAAGTG GGATGTATCT TTAAGTTTCT GGATGTACCT    506
GGACAGCTTC TTATCTCCGC GTATCCTTCA GCTTACCTAT GGACCTTTCC ATTCTATCTT    566
CAGTGATGAT GAACAATATC CCTATCTCTA TCAGATGGCC CCAAAGGACA CATCTCTAGC    626
ATTGGCAATG GTCTCCTTCA TACTTTATTT GAAATGGAAC TGGATTGGCC TTGTCATCCC    686
AGATGACGAT CAAGGAAACC AATTTCTTTT AGAGTTGAAG AAACAGAGTG AAAACAAGGA    746
AATTTGCTTT GCCTTTGTGA AAATGATCTC TGTTGATGAT ACTTCATTTC CACATAAAAC    806
TGAAATGGAC TACAACCAAA TTGTGATGTC ATCCACAAAT GTTATTATCA TTTATGGAGA    866
AACACGCAAT TTCATTTATT TGATCTTCAG AATGTGGGAA CCTCCCATTT TACAGAGAAT    926
ATGGATCACC ACAAAACAAT TGAATTTCCC TACCAGGAAG ACAGACATAA GTCATGGCAC    986
ATTCTATGGA TCACTTACTT TTCTACCCCA CCATGGTGAG ATTTCTGGCT TTAAAAAGTT   1046
TGTACAGACA TGGTTCCATG TCAGAAACAC AGATTTATAT TTAGTAATGC CAGAGTGGAA   1106
CTATTTTAAC TATGTAAGCT CAGCATCCAA TTGTAAAATA CTGAAGAACA ATTCATCTGA   1166
TGCCTCATTT GATTGGCTAA TGGAACAGAA GTTTGACATG ACCTTTAGTG AGAATAGTCA   1226
TAACATATAC AATGCTGTGC ATGCCATAGC CCATGCCCTC CATGAGATGA ATCTGCAACA   1286
GGCTGATAAT CAGGCAATAG GCAATGGAAA AGGAGCCAGT TCTCACTGCT TGAAGGTAAA   1346
CTCCTTTCTA AGAAGGACCT ACTTCACTAA TCCTCTTGGG GACAAAGTGT TTATGAAGCA   1406
AAGAGTAATA ATGCAGGATG AATATGATAT TATTCACTTT GGGAATCTCT CACAACACCT   1466
```

```
TGGGATTAAG ATGAAGTTAG GAAAGTTCAG CCCATATTTA CCACATGGTC GACACTCTCA   1526
CTTATATGTA GACATGATTG AGTTGGCCAC AGGAAGAAGA AAGATGCCAT CCTCTGTGTG   1586
CAGTGCAGAT TGTAGTCCTG GATTCAGAAG ATTGTGGAAG GAGGGAATGG CAGCCTGCTG   1646
TTTTGTTTGC AGCCCCTGCC CAGAAAATGA AATTTCTAAT GAGACAAGCT CCTCTCCATT   1706
TCATCCTTGC ATTCAGACAG GAACAATTAT GGGCTGGAGA TGTGACTATG GGATGGGAAT   1766
CCCATCACTC ACTTGATGTC CTGTCTTCCG GCTGGAGGTG GGCTCTTTAA GTTAACACTA   1826
TCTACTGTAG TACATTTCAT CTAAGGTCTC TGACCTCCCA AGTCTCTGGT GCATTTTGGT   1886
GGGTCCACCC ACCCTCCTAT TACCTGAAGT TGCCTGTTTA TATTCTTTTT GCTGGTCCTC   1946
AGAGATCGGT TCCCCTCTCA CCTGCCCACA CACCACAAAC CCCTTTCAAA TAACATCATA   2006
AATGATACAA TGAAGTTAAG TATACAAAGA ACAAATTGCT TGGTTTTATT TCATTTAAAT   2066
CTTTATGAAC TTTATGAATT GAAATCAATG CTCGGCAACA GCATCCTTCA CATTACATAT   2126
CAGCATCAAA GGCAGCATTG CAAGGCTTCT TTCATTACCC TTACTTGAAT TACCTTGACA   2186
ATAAAATTTC TGAAGCAGAC CTAACTAAGC TTTCCTTTGG AAATCAGATA TGGATCAATG   2246
TGTGAATTGT CCAGAATACC AATATGCCAA CACAGAACAG AACAAATGTA TTCAGAAAGG   2306
TGTCACCTTC CTAAGCTATG AAGACCCCTT GGGGATGGCA CTTGCCTTAA TGGCCTTCTG   2366
CTTCTCTGCA TTCACAGCTG TGGTACTTTG TGTCTTTGTG AAGCACCATG ACACTCCTAT   2426
TGTGAAGGCC AATAACAGAA GCCTCAGCTA CCTATTACTC ATGTCACTCA TGTTCTGTTT   2486
TCTGTGCTCC TTTTTCTTCA TTGGCCTTCC AAACAGAGCC ATCTGTGTCT TACAGCAAAT   2546
CACATTTGGA ATTGTATTCA CTGTGGCTGT TTCCACAGTT CTGGCCAAAA CAGTCACTGT   2606
GGTTCTGGCT TTCAAAGTCA CAGACCCAGG GAGAAGATTG AGAAACTTCC TGGTATCAGG   2666
GACACCCAAC TACATTATTC CCATATGTTC CCTACTCCAA TGTGTTCTGT GTGCAATCTG   2726
GCTAGCAGTT TCTCCTCCCT TTGTTGATAT TGATGAACAC ACTCTCCATG GCCATATCAT   2786
CATTGTGTGC AACAAGGGCT CAGATACTGC ATTCTACTGT ATCCTGGGAT ATTTGGCCTG   2846
CCTGGCACTT GGAAGCTTCT CTCTGGCTTT CTTGGCCAAG AATCTGCCTG ACACATTCAA   2906
TGAAGCCAAA TTCTTGACCT TCAGCATGCT AGTGTTCTGT AGTGTCTGGG TCACCTTCCT   2966
CCCTGTCTAC CATAGCACCA AGGGCAAACA CATGGTTGCT GTGGAGATCT TCTCCATCTT   3026
GGCATCCAGT GCAGGGATCC TTGGATGTAT TTTTGTACCC AAGATTTATA TCATTTTAAT   3086
GCGACCAGAG AGAAATTCTA CCCAAAAGAT CAGGGAAAAA TCATATTTCT GAACAAATAT   3146
TTAGGAATTC TGTCAAATGT AAAGTTGGTA CATACCCACC AAATATTGGG TTATAGTGCA   3206
TGTGTCTAGT TTTAGAATCA CTCTCACTGG TTGCTCTAGT GATATCAGGA AGTATCATAT   3266
CTACTGAACT TCCCTACAGT GTCCATAAAA TCTTGCACTC ATTCACTTTC TTCATTTTCT   3326
CTCAGAGAAC TAAACTCTCA ATTATTACAA TTTTATTCTT CATTTTGATT TCATGGAGAT   3386
GGCCCTCTGG TAACTGCCAA AAAATGTTGA TAAGGCAGTT GAATCCACCA CTTTGTGTAG   3446
AAAAATGAGA TCTAGGAAGA CAGGGTTACA CATAAAAACC ATCTACCAAA TCAAATAATC   3506
AATGAGAAAC ACAGACTAAC TAAATAATCA GCAAAGATGA AATCAGAACA TATTTTCTGA   3566
TTTCCAGTAA GAGCACACAC ATAAGAAAAT ACTTACTTTT TTCATCTGTT CTTCAATCTA   3626
CTGGCCAATA GTCTAAGGAG GAAATGTTCC TTTTCTGCTG TCAAATAAAA ATATATTATA   3686
TCC                                                                 3689
```

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 127 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```
Met Phe Ile Phe Met Gly Val Phe Phe Leu Leu Asn Ile Thr Leu Leu
 1               5                  10                  15
Met Ala Asn Phe Ile Asn Pro Arg Cys Phe Trp Arg Ile Asn Leu Asp
            20                  25                  30
Glu Ile Thr Asp Glu Tyr Leu Gly Leu Ser Cys Thr Phe Ile Leu Ala
        35                  40                  45
Ala Val Gln Thr Pro Thr Glu Lys Asp Tyr Phe Asn Lys Thr Leu Asn
    50                  55                  60
Val Leu Lys Thr Thr Lys Asn His Lys Tyr Ala Leu Ala Leu Val Phe
65                  70                  75                  80
Ala Met Asp Glu Ile Asn Arg Asn Pro Asp Leu Leu Pro Asn Met Ser
                85                  90                  95
Leu Ile Ile Arg Tyr Thr Leu Gly Leu Cys Asp Gly Lys Thr Val Thr
```

```
            100                 105                 110
Pro Thr Pro Tyr Leu Phe His Lys Lys Lys Thr Lys Pro Tyr Pro
        115                 120                 125
```

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3896 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 36...263
(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```
ATTTCACAAC TTCTTGATCT TAGACCTTAG CAGAT ATG AAA AAC CTG TGT GTT       53
                                       Met Lys Asn Leu Cys Val
                                        1               5

TTC ACT CTT TCC TTT TTC CTC CTG GAG TTT TCT CTG ATC TTG TGC CAT     101
Phe Thr Leu Ser Phe Phe Leu Leu Glu Phe Ser Leu Ile Leu Cys His
            10                  15                  20

TTG ACT GAA CCC ATT TGC TTT TGG AGG ATA AAT AAT AAT GAA GAT AAT     149
Leu Thr Glu Pro Ile Cys Phe Trp Arg Ile Asn Asn Asn Glu Asp Asn
        25                  30                  35

GAT GGA GAT TTG AGA AGT GAC TGT GGT TTT TTC CTT GCA GCA GTT GAG     197
Asp Gly Asp Leu Arg Ser Asp Cys Gly Phe Phe Leu Ala Ala Val Glu
    40                  45                  50

GGA CCT ACT GAC GAC TCT TAT AAT ATC TCT GAT CTT AGG TTT TCT TTG     245
Gly Pro Thr Asp Asp Ser Tyr Asn Ile Ser Asp Leu Arg Phe Ser Leu
55                  60                  65                  70

GAC CAT TTA ATC CTA AGC TGAGTGACCA TGACCAGTTT CCCTATGTCC ATCAGGTA   301
Asp His Leu Ile Leu Ser
                75

GCCACCAAGG ACACACGTTT GTCCCATGCA ATGGTCTCCT TGATGTTTCA TTTTACATGG   361
ATTTGGATAG GAATGGTCAT CTCAGATGAT GACCAGAGTA TTCAGTTTCT ATCAGACATG   421
AGAGAAGAAA TGCAAAGACA TGGAATCTGT TTAGCTTTTG TTAATATGAT CCCAGAAGAC   481
ATGCAGTTAT ATATGACAAG GGCTACAATA TATGATAAAC AAATTATGGA ATCAACAGCA   541
AAGGTTGTTA TGATTTATGG TGAAATGAAC TCTACCTTAG AAGTTAGCTT TAGAAGGTGG   601
GAAGATTTAA GTATAAGGAG AATCTGGATC ACAACCTCAC AATGGGACGT TATCACAAAT   661
AAAAATGATT TCAGCCTTGA TTTCTTCCAA GGGACTGTCA CTTTTGCACA CCATGTAGGT   721
GAAATTGCTA ACTTTAGGAA TTTCTTGCAA ACAATGAACA GTGAAAAATA CACAGTAAAC   781
ATTTCTGAGT CTAGACTGGG GTGGAATTAT TTTAATTGTT CCATCTCTAA GAACAGCAAT   841
AAAAAGGATC ATTTTACATT CAACAACACA TTGGAATGGA CAACACTGCA CAAATATGAC   901
ATGGTCCTAA GTGAGGAAGG CTACAATTTG TATAATGCTG TGTATGCTGT GGCCCACACC   961
TACCATGAAC TCGTTCTTCA ACAAGTAGAA TCTCAGCAAA TGACAGTACC CAAAGGAACA  1021
TTCACTGACT GTCAGCAGGT GTCTTCCATG CTGAAGTCCA GGATATTTAC TAACCCTGTT  1081
GGAGAACTGG TGAACATGAA GCATAGGGAA AATCAGTGTA CAGAGTATGA TATTTTCATC  1141
ATTTGGAATT TTCCACAAGG CCTTGGATTA AAAGTGAAAA TAGGAAGCTA TTTGCCTTGT  1201
TTCCAACAGA GCCAACAACT TCATATATCT GAAGATTTGG AGTGGGCCAC AGGAGGATCA  1261
TCAGTACCCC CCTCCCTGTG TAGTGTAACA TGTACTGCTG GATTCAGGAA AATTCATCAG  1321
AAACAAACAG CAGACTGCTG CTTTGATTGT GATCAGTGCC CAGAAAATGC AGTTTCCAAT  1381
GAAACAGAGA TATGCAATCT GAACATGGAA AGACCATCAT TATTTGCAAC AAAGGCTCAG  1441
```

```
TAATTGCCTT CCACTTTGTT CTCGGATACT TGGGTGCCTT GGCTCTGGGG AGCTTTACTG   1501
TGGCTTTCTT GGCTAGGAAC CTTCCTGACA GATTCAATGA AGCCAAATTC TTAACCTTCA   1561
GCATGCTGGT GTTCTGCAGT GTCTGGATCA CCTTCCTCCC TGTCTACCAC AGCACCCAGG   1621
GAACGGTCAT GGTGGTTGTG GAGGTTTTCT CCATCTTGGC TTCTAGTGCA GGCTTGCTAG   1681
GGTGTATCTT TCTCCCAAAA TGTTGTGTTT TATTACGTAT ACAAAATTCA AACTTTCTGC   1741
ATAAGTACAA ACATGAATTG CATTCTTGAT TCTTTAGTAA TTTAAAAATG CTAATCATAC   1801
TCAACTTATC TTTTTGCTTT GTCATAACAA AAGCACCACT AAATCATACA AAAAATTTAA   1861
GTAATATACA AATTTAGTAT TTACAATGTA GGGCAGCACA GCACTGCCTA ATGTAATGCC   1921
AATTATTGTT TTAGAGGTAA ATGGTCTTAT TCATGTGTAC ATAGATGTAA ACATTGAGAA   1981
TAGGGAATCT AACTTGATGA ATGGCTATCA ACACTTTGAC CTCTAGGTAT GTGTGTAAGC   2041
CATGTACCTA ATTTAATATG TAATAAGGTG AGCGTAACAT ATGTGAGAGT GCTACCTCTG   2101
GGCAGAAAGT TCTGGGAATT ATAAGAAAGA GGACTTCAAA GAGCACAGGC ATGAAGTCAA   2161
TAATCAGCAT TATTCCATGT GCTCTCATTG AGTGTCTGCA TCCACGTTCT TGTCTTGACT   2221
TCATTCTATT AACTGTGACT AAGGTACATA GGGAAATAGG ACTTTTCTCA CATGGTTCCT   2281
TTGACCATGG TGTTTTCTTA CAGCAACAGA CTCTAAGACA TCAGCAAAAT GTTAAATTGC   2341
CTTGGTTAGG ATTTGGAATA TCACAGATTA CTGATGCAAT AGAAGGCACT GATTTGAAAG   2401
AGAAAATAGA TTGAATACTA GGGGAGTGTG AGCATAGTTA CAGTGTTGCA TATTGTTGAT   2461
GGCCATCACA GAGGCCTGAG ATTTGTAATT GCTTCATAAT GTACTATGAA AATATTCAGA   2521
ATATCAGGTA ACATACTAAA AGAAGTACAA TATATGAAAA GGACAATGGG GTTCAGATTA   2581
TGCCTGCTCT ATAAGGCTCA TGAACTTCAT ATGAAAACAT ACCATTTCAA TATGAAATGA   2641
AGAAGTTTCA TTCAGGGAGA AAAATTGGTA GTGGAAAAAT TTACACACAA GACCTATATC   2701
ACAAGGAGAT CAGTGAAATC TTGGAATATA TAAGGCACTC TAGAAGAATG ACTTCAAAAA   2761
TGTTAGCAAA ATAGGAACAA CTAAGAATTA TTTGGTTTAA TATTACATAA TCAAAGATGT   2821
ACATACAAAC ACATGAACAT TATTATTTCT GGACGTCAGT TGCTGAAGGT CAGTGTCATT   2881
TTCTCTCAAA GTATTGTTTG TTGCTCTTAT TTTACTTGTT AATTTACAGT TTATTTTTGA   2941
TGGGATAATT TAATTGTTTT TTTCTTTATA TTTCCTGTCT CAAGAACACC ACTTGTAGCC   3001
CATCCATACA CTCCTAAAAT GCAAATGACC TATTATTTCA TTAATGCTTA ATGAATGCAT   3061
GCATGTATTT GTATATACAT ATACATTTTA AAGTATACAT TGTAGATACT ATGTAAAATT   3121
GCATGTTTTT ATGTTTTGAT GGCTCATTAT TTGGTAATAC CTGGCCAATA TTTGTTCCCT   3181
TCCCTGGCTA TGACAACCTC CTCCATTCCC TGATTTAAAG TTTCCTGTAA ATGGTTGTGT   3241
AGGGTAGAAG CTTTGAAAGC TTTCTTCCTT CCACGCTGCC ATGCACAGTG CAGTAATCCT   3301
TCTTCAGACC ATATTTTGTG TGTCATATTG GTAAAACTTC ATGGTCTACT TATGCTAGTT   3361
CTAGAAGATT TGTGTTCACA GCCAGTTTCC TCATCCTTTG ACTCACAAGA TCTTTTCCAC   3421
CATCTTCTTT ACGTTTCTCT GAGCCTTGGA TGAGGGAAAA TTTTGTAAGA GGATACATTG   3481
AATTGTTTCC TTCAACTACC TACTCTGGAA ATGACTATCA CACTATCACA ACATCTTTAA   3541
AAACAAGATG GAACTCCAAA ATCATTTTCT AAGGAAATAA ATGAAAATCT AAGTGTTCTT   3601
TTAATCTGGT TCATTGGAAT TTCCTGCATT TATCTGCCTG GGTGTATGTA ATCCCCCCCC   3661
CCCAGCCTGA AACCTGGCTG AACAGGTTTC ACTGTTAGCA CGAAGAGAGA ATCCGGGGTG   3721
GAGCCTTCCA CCCTATCATT CTGCCACTCC CACTGCTACT GCCTGCCGCC CAGCTGTTCC   3781
GGAGCTATCA CGTGGTCACC TGAAATTGGA CTCCAAGGAT GATTTGGAGG GAATGGGTGC   3841
CTTCCCCTTC TTCATAAACC AGTGTCTGGG AATAGTAAAA TTGAACTTTG ATCAG        3896
```

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 76 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

```
Met Lys Asn Leu Cys Val Phe Thr Leu Ser Phe Phe Leu Leu Glu Phe
 1               5                  10                  15
Ser Leu Ile Leu Cys His Leu Thr Glu Pro Ile Cys Phe Trp Arg Ile
            20                  25                  30
Asn Asn Asn Glu Asp Asn Asp Gly Asp Leu Arg Ser Asp Cys Gly Phe
        35                  40                  45
Phe Leu Ala Ala Val Glu Gly Pro Thr Asp Asp Ser Tyr Asn Ile Ser
    50                  55                  60
Asp Leu Arg Phe Ser Leu Asp His Leu Ile Leu Ser
65                  70                  75
```

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2811 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 962...2605
(D) OTHER INFORMATION: GoVN1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```
GAAACGTCTA CTAATATGCT GTTCTCTTGG CTTTTTATCT CCTTGTTTCT ACAGATGCCA     60
ACTCTCATCT GGACCATTGC AACCCCTTCC TGCCTAACTG AATCAGGATA CCTCGTACAC    120
CAGGATGGAG CTGTGGTCAT TGGTGCATTT TTTCCTGTTT TAAAGTCCTT GCCTATAAGT    180
GAAATAATAG ATTGGAAAAC ATTATCTTTT GACACATACA ATTCTTTATG GATAAATGCA    240
CAAATGTACC AACTTGTTTT GGCCATGATA TTTGCGATCA ATGAGATCAA TGTGAAGTCC    300
CATATTTTAC CAAATACCTC TCTGGGACTT GAGATTTATA ATCTGCCATA TTTTGAACGG    360
AATATTCTGA GGAGTGCACT ATCTTGGCTC ACAGGCTTGA GCAAATTCAT TCCTAATTAC    420
ACCTGCAGAA AGGATAGCAA ATCAGCTGCT GCACTTACTG GAATATCACA GAAAACATCT    480
GAGACCTTTG GGACTTTGTT GGACATTTAC AAATTTCCTC AGCTTAATTT TGGGCCGTGT    540
GATCCTGTTC AGATAGGCAG AAACCAGTTT CCATCTGTGT ACCAGGTGGC CCCCAAAGAC    600
ACACCTCTGT TCTGTGGTAT CACCTCTTTG ATGCTTCATT TCAACTGGAC CTGGGTGGGA    660
CTGCTAATCA CAGATGACAA CAGAGGTTCT CAGTTTCTAT CAGAGTTAAG AAAGGAGCTG    720
GACAAGAATA AAATCTGCAT AGCCTTTGTG GAAACAGTAA TATTTTTTGG GGAATCATTG    780
CATTATATGC TAACCCACAA TCAGATGCAG ACTCTAGAGT CATCAGCAAA TGTGATTATA    840
GTTTATGGAC ATTTTGCTTT TCAATTAATT GTAATACAAA GTAAACACAG AAAGTATGAA    900
ATGAAAAAGA TTTGGGTCAT AACCTCAAAA TGGGTTGGCC AAAAAAATTG AACAATATAC    960
C ATG TTA GAA TTG GCC CAT GGC ACT CTG ACT TTC TCA CCC CAT CAT GGG   1009
  Met Leu Glu Leu Ala His Gly Thr Leu Thr Phe Ser Pro His His Gly
   1               5                  10                  15

GAG ATT TCT GAT TTC ACA AAT TTT ATG CAG GAA GTC ACC CCT ATC AAG    1057
Glu Ile Ser Asp Phe Thr Asn Phe Met Gln Glu Val Thr Pro Ile Lys
            20                  25                  30

TAC CCA GAA GAC ATT TTT CTT CAC ATC TTG TGG AAC CAG TAT TTC AAT    1105
Tyr Pro Glu Asp Ile Phe Leu His Ile Leu Trp Asn Gln Tyr Phe Asn
        35                  40                  45

TGT CCA CTT TTG CAT TCT GAG TGT AAA ATC TTT GAA AAC TGT ATA CCC    1153
Cys Pro Leu Leu His Ser Glu Cys Lys Ile Phe Glu Asn Cys Ile Pro
    50                  55                  60

AAT GCC TCT TTG GAA TTG TTG CCA GGG GGT GTT TTT GAG CTG GTC ATG    1201
Asn Ala Ser Leu Glu Leu Leu Pro Gly Gly Val Phe Glu Leu Val Met
65                  70                  75                  80

ACT GAA GAG AGT TAC AAT GTG TAC AAT GCT GTG TAT GCA GTG GCC CAC    1249
Thr Glu Glu Ser Tyr Asn Val Tyr Asn Ala Val Tyr Ala Val Ala His
                85                  90                  95

AGT CTC CAT GAG AAG GCT CTC CAT CAA GTA GAA ATT CAA CCA CAG GAT    1297
Ser Leu His Glu Lys Ala Leu His Gln Val Glu Ile Gln Pro Gln Asp
            100                 105                 110

AAT AAA GAT AGG ACT ATA TTA TTT CCT TGG CAG CTT CAC CCT TTT CTG    1345
Asn Lys Asp Arg Thr Ile Leu Phe Pro Trp Gln Leu His Pro Phe Leu
        115                 120                 125
```

```
AAG AAC ATT CAG CTG ATA AAT TCT GTT GGT GAT CGT GTG ATT CTG GAC    1393
Lys Asn Ile Gln Leu Ile Asn Ser Val Gly Asp Arg Val Ile Leu Asp
    130                 135                 140

TGG AAA AAG AAG ACG GAT ACA GAG TAT GAT ATT TCC AAT ATT TGG AAT    1441
Trp Lys Lys Lys Thr Asp Thr Glu Tyr Asp Ile Ser Asn Ile Trp Asn
145                 150                 155                 160

TTC CCA ACA GGT CTT TCC TTA TTA GTG AAA GTG GGT ACA TTT GCT CCA    1489
Phe Pro Thr Gly Leu Ser Leu Leu Val Lys Val Gly Thr Phe Ala Pro
                165                 170                 175

AGT GCT CCC AAG GGG GAA CAA CTT TCG ATA TCT GAA CAC ACA ATT AAC    1537
Ser Ala Pro Lys Gly Glu Gln Leu Ser Ile Ser Glu His Thr Ile Asn
            180                 185                 190

TGG CCC ATA GGA TTT ACA GAG ATT CCA AAG TCT GTA TGC AGT GAG AGC    1585
Trp Pro Ile Gly Phe Thr Glu Ile Pro Lys Ser Val Cys Ser Glu Ser
        195                 200                 205

TGC AGT CCT GGA CAC AGG AAA GTC ATC CTG GAG AGC AAG CCT GCC TGT    1633
Cys Ser Pro Gly His Arg Lys Val Ile Leu Glu Ser Lys Pro Ala Cys
    210                 215                 220

TGC TTT GAC TGC ACT CCT TGC CCA GAT AAA GAG ATT TCC AAC GAG ACA    1681
Cys Phe Asp Cys Thr Pro Cys Pro Asp Lys Glu Ile Ser Asn Glu Thr
225                 230                 235                 240

GAT GTG GGT CAG TGT GTG AAG TGT CCT GAA TCT CAT TAT GCA AAT ACA    1729
Asp Val Gly Gln Cys Val Lys Cys Pro Glu Ser His Tyr Ala Asn Thr
                245                 250                 255

GAG AAG AGT CAC TGC CTG AAG AAG ACT ATG ACC TTT CTG GAT TAT AAT    1777
Glu Lys Ser His Cys Leu Lys Lys Thr Met Thr Phe Leu Asp Tyr Asn
            260                 265                 270

GAT TCC TTG GGG ACG GGA CTC ACA CTC ATG TCT CTG GGA TTC TTT GTT    1825
Asp Ser Leu Gly Thr Gly Leu Thr Leu Met Ser Leu Gly Phe Phe Val
        275                 280                 285

GTC ACA GGT CTT GTT ATT GGG GTT TTT ATA ATC CAC AGA AAC ACT CCA    1873
Val Thr Gly Leu Val Ile Gly Val Phe Ile Ile His Arg Asn Thr Pro
    290                 295                 300

ATT GTG AAG GCC AAT AAT AGA TCT CTC AGT TAT ATC CTG CTC ATC ACT    1921
Ile Val Lys Ala Asn Asn Arg Ser Leu Ser Tyr Ile Leu Leu Ile Thr
305                 310                 315                 320

CTC ACT CTC TGT TTC CTT TGT CCC TTG CTC TTC ATT GGG CTT CCA AAC    1969
Leu Thr Leu Cys Phe Leu Cys Pro Leu Leu Phe Ile Gly Leu Pro Asn
                325                 330                 335

ACA GCC ACA TGT ATC CTA CAG CAG AAC TTG TTT GGA CTT CTC TTC ACT    2017
Thr Ala Thr Cys Ile Leu Gln Gln Asn Leu Phe Gly Leu Leu Phe Thr
            340                 345                 350

GTG GCT CTA TCC ACA GTG TTG GCC AAA ACT ATC ACT GTA GTT ATG GCA    2065
Val Ala Leu Ser Thr Val Leu Ala Lys Thr Ile Thr Val Val Met Ala
        355                 360                 365

TTC AAG ATT ACT GCT CCA GGA AGA AAG ACA AGA TGG TTG CTG ATA TTA    2113
Phe Lys Ile Thr Ala Pro Gly Arg Lys Thr Arg Trp Leu Leu Ile Leu
    370                 375                 380

AGA GCC CCT CAG TTC ATC ATT CCA CTT TGT GCC CTG ATG CAA ATC CTT    2161
```

```
Arg Ala Pro Gln Phe Ile Ile Pro Leu Cys Ala Leu Met Gln Ile Leu
385                 390                 395                 400

TTC TCT GGG ATA TGG CTG GGA ACA TCT CCT CCA TTT GTT GAC ATG GAT    2209
Phe Ser Gly Ile Trp Leu Gly Thr Ser Pro Pro Phe Val Asp Met Asp
                405                 410                 415

GCT CAC TCT GAA CAT GGG CAC ATC ATC ATT CTA TGC AAC AAG GGC TCA    2257
Ala His Ser Glu His Gly His Ile Ile Ile Leu Cys Asn Lys Gly Ser
            420                 425                 430

GCT ATT GGC TTC TAC TGT ACT CTG GCC TAC CTG GGA GTC ATG GCC TTT    2305
Ala Ile Gly Phe Tyr Cys Thr Leu Ala Tyr Leu Gly Val Met Ala Phe
        435                 440                 445

GGT AGT TAC CTC TTG GCT TTC ATG TCC AGG AAT CTT CCT GAC ACA TTT    2353
Gly Ser Tyr Leu Leu Ala Phe Met Ser Arg Asn Leu Pro Asp Thr Phe
    450                 455                 460

AAT GAA TCC AAG GCC CTG GCT TTC AGC ATG CTG ATG TTC TGC AGT GTC    2401
Asn Glu Ser Lys Ala Leu Ala Phe Ser Met Leu Met Phe Cys Ser Val
465                 470                 475                 480

TGG GTC ACA TTC CTC CCT GTC TAC CAC AGC ACC ACT GGG AAG GTC AGG    2449
Trp Val Thr Phe Leu Pro Val Tyr His Ser Thr Thr Gly Lys Val Arg
                485                 490                 495

GTG GCT ATG GAA ATG TTT TCT ATC TTG GCT TCC AGT GCA AGC ATT CTA    2497
Val Ala Met Glu Met Phe Ser Ile Leu Ala Ser Ser Ala Ser Ile Leu
            500                 505                 510

ACC CTA ATC TTT GTC CCT AAG TGC TAC ATT GTT TTG TTC AGA CCA GAG    2545
Thr Leu Ile Phe Val Pro Lys Cys Tyr Ile Val Leu Phe Arg Pro Glu
        515                 520                 525

AGG AAC ATA CTT CCT CTA AAC AGA GAA AAA AGA CAG CAT AGG AGT AAA    2593
Arg Asn Ile Leu Pro Leu Asn Arg Glu Lys Arg Gln His Arg Ser Lys
    530                 535                 540

AAT TCT GAA ACA TAGCAGTCAA GACAAACATT GGCCTAGCAC AAAATGTCTG ATTGT  2650
Asn Ser Glu Thr
545

TGGCATTTCT CCTGCTATAT AAACAATTAG TCCTTTGACT TTGAGGACAG GATCACATGA   2710
GACAGACCGG TGATATTGCT TCAAATTATG TAAAATATGT GACATGGTTA TATTGACCAA   2770
TAAAATACTT GTTCTTGTAT GAAAAAAAAA AAAAAAAAAA A                       2811
```

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 548 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```
Met Leu Glu Leu Ala His Gly Thr Leu Thr Phe Ser Pro His His Gly
 1               5                  10                  15
Glu Ile Ser Asp Phe Thr Asn Phe Met Gln Glu Val Thr Pro Ile Lys
            20                  25                  30
Tyr Pro Glu Asp Ile Phe Leu His Ile Leu Trp Asn Gln Tyr Phe Asn
```

```
          35                  40                  45
Cys Pro Leu Leu His Ser Glu Cys Lys Ile Phe Glu Asn Cys Ile Pro
    50                  55                  60
Asn Ala Ser Leu Glu Leu Leu Pro Gly Gly Val Phe Glu Leu Val Met
65                  70                  75                  80
Thr Glu Glu Ser Tyr Asn Val Tyr Asn Ala Val Tyr Ala Val Ala His
                85                  90                  95
Ser Leu His Glu Lys Ala Leu His Gln Val Glu Ile Gln Pro Gln Asp
            100                 105                 110
Asn Lys Asp Arg Thr Ile Leu Phe Pro Trp Gln Leu His Pro Phe Leu
        115                 120                 125
Lys Asn Ile Gln Leu Ile Asn Ser Val Gly Asp Arg Val Ile Leu Asp
    130                 135                 140
Trp Lys Lys Lys Thr Asp Thr Glu Tyr Asp Ile Ser Asn Ile Trp Asn
145                 150                 155                 160
Phe Pro Thr Gly Leu Ser Leu Leu Val Lys Val Gly Thr Phe Ala Pro
                165                 170                 175
Ser Ala Pro Lys Gly Glu Gln Leu Ser Ile Ser Glu His Thr Ile Asn
            180                 185                 190
Trp Pro Ile Gly Phe Thr Glu Ile Pro Lys Ser Val Cys Ser Glu Ser
        195                 200                 205
Cys Ser Pro Gly His Arg Lys Val Ile Leu Glu Ser Lys Pro Ala Cys
    210                 215                 220
Cys Phe Asp Cys Thr Pro Cys Pro Asp Lys Glu Ile Ser Asn Glu Thr
225                 230                 235                 240
Asp Val Gly Gln Cys Val Lys Cys Pro Glu Ser His Tyr Ala Asn Thr
                245                 250                 255
Glu Lys Ser His Cys Leu Lys Lys Thr Met Thr Phe Leu Asp Tyr Asn
            260                 265                 270
Asp Ser Leu Gly Thr Gly Leu Thr Leu Met Ser Leu Gly Phe Phe Val
        275                 280                 285
Val Thr Gly Leu Val Ile Gly Val Phe Ile Ile His Arg Asn Thr Pro
    290                 295                 300
Ile Val Lys Ala Asn Asn Arg Ser Leu Ser Tyr Ile Leu Leu Ile Thr
305                 310                 315                 320
Leu Thr Leu Cys Phe Leu Cys Pro Leu Leu Phe Ile Gly Leu Pro Asn
                325                 330                 335
Thr Ala Thr Cys Ile Leu Gln Gln Asn Leu Phe Gly Leu Leu Phe Thr
            340                 345                 350
Val Ala Leu Ser Thr Val Leu Ala Lys Thr Ile Thr Val Val Met Ala
        355                 360                 365
Phe Lys Ile Thr Ala Pro Gly Arg Lys Thr Arg Trp Leu Leu Ile Leu
    370                 375                 380
Arg Ala Pro Gln Phe Ile Ile Pro Leu Cys Ala Leu Met Gln Ile Leu
385                 390                 395                 400
Phe Ser Gly Ile Trp Leu Gly Thr Ser Pro Pro Phe Val Asp Met Asp
                405                 410                 415
Ala His Ser Glu His Gly His Ile Ile Ile Leu Cys Asn Lys Gly Ser
            420                 425                 430
Ala Ile Gly Phe Tyr Cys Thr Leu Ala Tyr Leu Gly Val Met Ala Phe
        435                 440                 445
Gly Ser Tyr Leu Leu Ala Phe Met Ser Arg Asn Leu Pro Asp Thr Phe
    450                 455                 460
Asn Glu Ser Lys Ala Leu Ala Phe Ser Met Leu Met Phe Cys Ser Val
465                 470                 475                 480
Trp Val Thr Phe Leu Pro Val Tyr His Ser Thr Thr Gly Lys Val Arg
                485                 490                 495
Val Ala Met Glu Met Phe Ser Ile Leu Ala Ser Ser Ala Ser Ile Leu
            500                 505                 510
Thr Leu Ile Phe Val Pro Lys Cys Tyr Ile Val Leu Phe Arg Pro Glu
        515                 520                 525
Arg Asn Ile Leu Pro Leu Asn Arg Glu Lys Arg Gln His Arg Ser Lys
    530                 535                 540
Asn Ser Glu Thr
545
```

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3584 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 273...2576
(D) OTHER INFORMATION: GoVN2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```
CACACTGCCC AGGTTTAAGG CAGAAAGAAT ATGTTCATTT TGATGGTAGT ATTTTTCCTT     60
CTCCACCATC CACTTCTCAT GGCAAATTTC ATCGATCCCT GGTGCTTTTG GAGAACAAAT    120
TTGAATGAAG TCAAGGAAAA AAACTTGGAT ATAAATTGTG CCTTCATCCT TGGAGCAGTT    180
CAGTTGCCTA TGGAGAAAGA TATTTCAATG AGACTTTGAA TGTCCTAAAA ACAACTAAAA    240
ACAACAAATA TGCCTTGGCA TTAGCCTTTT CA ATG GAG GAA ATC AAC AGG AAC      293
                                   Met Glu Glu Ile Asn Arg Asn
                                    1               5

CCT GAT CTT TTA CCA AAT ATG TCT TTG GTT ATA AAA CAT ACT TTG AGC      341
Pro Asp Leu Leu Pro Asn Met Ser Leu Val Ile Lys His Thr Leu Ser
        10                  15                  20

TAT TGT GAT GGA AAT ACT GCA GAC CAT ATA TTT AAA GAA AAA TTT TAT      389
Tyr Cys Asp Gly Asn Thr Ala Asp His Ile Phe Lys Glu Lys Phe Tyr
    25                  30                  35

AAG CCT TTA CCT AAT TAT GTC TGT AAT GAA GAG ACT ATG TGT TCA TTT      437
Lys Pro Leu Pro Asn Tyr Val Cys Asn Glu Glu Thr Met Cys Ser Phe
40                  45                  50                  55

ATG CTT ATA GGG CTG AAT TGG GTA TTG TCT CTA ACA CTT TTT AAA GAC      485
Met Leu Ile Gly Leu Asn Trp Val Leu Ser Leu Thr Leu Phe Lys Asp
                60                  65                  70

TTG GAC ATC TTC TCA TTT CCA CGT TTC CTT CAA ATT TCC TAT GGA CCT      533
Leu Asp Ile Phe Ser Phe Pro Arg Phe Leu Gln Ile Ser Tyr Gly Pro
            75                  80                  85

TTC CAT TCC ATC TTC AGT GAT AAT GAA CAA TTT CCA TAT CTC TAT CAG      581
Phe His Ser Ile Phe Ser Asp Asn Glu Gln Phe Pro Tyr Leu Tyr Gln
        90                  95                  100

ATG ACC CCA AAG GAC ACA TCA CTA GCA TTG GCA ATT GTC TCC TTC TTA      629
Met Thr Pro Lys Asp Thr Ser Leu Ala Leu Ala Ile Val Ser Phe Leu
    105                 110                 115

CTT TAC TTC AAT TGG AAC TGG GTT GGG CTT GTC ATC TCT GAT AAT GAT      677
Leu Tyr Phe Asn Trp Asn Trp Val Gly Leu Val Ile Ser Asp Asn Asp
120                 125                 130                 135

GAA GGC AAT CAA TTT CTC TCA GAG TTG AAA AAA GAG ACC CAA AAC AAG      725
Glu Gly Asn Gln Phe Leu Ser Glu Leu Lys Lys Glu Thr Gln Asn Lys
                140                 145                 150

GAA ATT TGC TTT GCC TTT GTT AAC ATG ATG TCA ATC CAT GAG CAT TCA      773
Glu Ile Cys Phe Ala Phe Val Asn Met Met Ser Ile His Glu His Ser
            155                 160                 165
```

```
TCT TAT CAA AAA ACT GAA ATG TAC TAC AAT CAA ATA GTG ATG TCA TCA      821
Ser Tyr Gln Lys Thr Glu Met Tyr Tyr Asn Gln Ile Val Met Ser Ser
        170                 175                 180

ACA AAT ATT ATT ATC ATT TAT GGG AAA ACA AAC AGT ATC ATT GAA TTG      869
Thr Asn Ile Ile Ile Ile Tyr Gly Lys Thr Asn Ser Ile Ile Glu Leu
    185                 190                 195

AGC TTC AGA ATG TGG GTA TCT CCA GTT ATA CAG AGG ATT TGG GTC ACA      917
Ser Phe Arg Met Trp Val Ser Pro Val Ile Gln Arg Ile Trp Val Thr
200                 205                 210                 215

AAC TCA GAG TTG GAT TTC CCG ACA AGT ATG AGA GAC TTC ACT CAT GGC      965
Asn Ser Glu Leu Asp Phe Pro Thr Ser Met Arg Asp Phe Thr His Gly
                220                 225                 230

ACA TTC TAT GGG ACT CTG ACA TTT CTA CAC CAC CAT GGT GAG ATT TCT     1013
Thr Phe Tyr Gly Thr Leu Thr Phe Leu His His His Gly Glu Ile Ser
            235                 240                 245

GGA TTT ACA AAT TTT TTC GAG ACA TGG GAC CAT CTC AGA AGC AGA GAT     1061
Gly Phe Thr Asn Phe Phe Glu Thr Trp Asp His Leu Arg Ser Arg Asp
        250                 255                 260

TTA AAT CTA TTA ATA CCA GAG TGG AAG TAC TTT AGC TAT GAT GCC TCA     1109
Leu Asn Leu Leu Ile Pro Glu Trp Lys Tyr Phe Ser Tyr Asp Ala Ser
    265                 270                 275

GGA TCT AAC TGT AAA ATA TTG AGG AAC TAT TCA TCC AAT GCC TCA TTG     1157
Gly Ser Asn Cys Lys Ile Leu Arg Asn Tyr Ser Ser Asn Ala Ser Leu
280                 285                 290                 295

GAA TGG ATA ACA GAA CAG AAG TTT CAC ATG GCC TTT AAT GAT TAT AGT     1205
Glu Trp Ile Thr Glu Gln Lys Phe His Met Ala Phe Asn Asp Tyr Ser
                300                 305                 310

CAT AGT ATA TAT AAT GCT GTG TAT GCC ATG GCC CAT GCC CTC CAT GAG     1253
His Ser Ile Tyr Asn Ala Val Tyr Ala Met Ala His Ala Leu His Glu
            315                 320                 325

ACT AAT CTG CAA GAG GTT GAT AAT AAG GAA ATA AGA AAT GGG AAA GGA     1301
Thr Asn Leu Gln Glu Val Asp Asn Lys Glu Ile Arg Asn Gly Lys Gly
        330                 335                 340

GCA AGT ACT CAC TGC TTG AAG GTA AAC TCA TTT CTC AGA AAG ACC CAC     1349
Ala Ser Thr His Cys Leu Lys Val Asn Ser Phe Leu Arg Lys Thr His
    345                 350                 355

TTT ACT AAT TCT CAT GGA GAG AGA GTG ATT ATG AAA CAG AGA GTG AGA     1397
Phe Thr Asn Ser His Gly Glu Arg Val Ile Met Lys Gln Arg Val Arg
360                 365                 370                 375

GTA CAG GAA GAC TAT GAC ATT GTT CAC ATT CAG AAT TTC TCA CAA CAC     1445
Val Gln Glu Asp Tyr Asp Ile Val His Ile Gln Asn Phe Ser Gln His
                380                 385                 390

CTT CGG ATT AAG ATG AAG ATA GGA AAG TTC AGC CCA TAT TTT ACA CAT     1493
Leu Arg Ile Lys Met Lys Ile Gly Lys Phe Ser Pro Tyr Phe Thr His
            395                 400                 405

GGT GGA CCC TTT CAC TTA TAT GAA GAC ATG ATT CAG TTG GCC ACA GGA     1541
Gly Gly Pro Phe His Leu Tyr Glu Asp Met Ile Gln Leu Ala Thr Gly
        410                 415                 420

AGT AGA AAG ATG CCG TCC TCT GTG TGC AGT GCA GAT TGT AGT CCT GGA     1589
```

```
Ser Arg Lys Met Pro Ser Ser Val Cys Ser Ala Asp Cys Ser Pro Gly
    425                 430                 435

TTC AGA AAA TCC TGG AAG GAG GGA ATG GCC CCC TGC TGT TTT ATT TGC    1637
Phe Arg Lys Ser Trp Lys Glu Gly Met Ala Pro Cys Cys Phe Ile Cys
440                 445                 450                 455

AGC CTG TGC CCT GAA AAT GAA ATT TCT AAT GAG ACA AAT ATG GAT CAA    1685
Ser Leu Cys Pro Glu Asn Glu Ile Ser Asn Glu Thr Asn Met Asp Gln
                460                 465                 470

TGT GTG AAT TGT CCA GAA TAC CAA TAT GCC AAC ACA GAA AAG AAC AAA    1733
Cys Val Asn Cys Pro Glu Tyr Gln Tyr Ala Asn Thr Glu Lys Asn Lys
            475                 480                 485

TGC ATT CAG AAA GAC GTG ATT TTT CTA AGC TAT GAA GAC CCC TTG GGA    1781
Cys Ile Gln Lys Asp Val Ile Phe Leu Ser Tyr Glu Asp Pro Leu Gly
        490                 495                 500

ATG GCT CTT GCC TTA ATT GCC TTC TGT TTG TCT GCA TTC ACA GCT GTG    1829
Met Ala Leu Ala Leu Ile Ala Phe Cys Leu Ser Ala Phe Thr Ala Val
    505                 510                 515

GTA CTT TGG GTC TTT GTG AAG CAC CAT GAC ACT CCT ATT GTG AAG GCC    1877
Val Leu Trp Val Phe Val Lys His His Asp Thr Pro Ile Val Lys Ala
520                 525                 530                 535

AAT AAC AGA ATC CTC AGC TAC ATA TTA ATC ATG TCA CTA ATG TTC TGT    1925
Asn Asn Arg Ile Leu Ser Tyr Ile Leu Ile Met Ser Leu Met Phe Cys
                540                 545                 550

TTT CTC TGC TCC TTT TTC TTC ATT GGC CAT CCT AAC AGA GGT ACC TGT    1973
Phe Leu Cys Ser Phe Phe Phe Ile Gly His Pro Asn Arg Gly Thr Cys
            555                 560                 565

ATC TTA CAG CAA ATC ACA TTT GGC ATT GTA TTC ACT GTG GCT GTT TCC    2021
Ile Leu Gln Gln Ile Thr Phe Gly Ile Val Phe Thr Val Ala Val Ser
        570                 575                 580

ACA GTT CTG GCC AAA ACA ATC ACT GTC ATT CTT GCT TTC AAA CTC AGA    2069
Thr Val Leu Ala Lys Thr Ile Thr Val Ile Leu Ala Phe Lys Leu Arg
    585                 590                 595

GAC CCA GGG AGA AGT TTA AGA AAC TTC CTG GTA TCT GGT GCA CCC AAC    2117
Asp Pro Gly Arg Ser Leu Arg Asn Phe Leu Val Ser Gly Ala Pro Asn
600                 605                 610                 615

TAC ATT ATT CCT ATA TGT TCC TTA TTG CAA TGT ATT CTG TGT GCA ATT    2165
Tyr Ile Ile Pro Ile Cys Ser Leu Leu Gln Cys Ile Leu Cys Ala Ile
                620                 625                 630

TGG CTA GCA GTT TCT CCT CCT TTT GTT GAT ATT GAT GAA CAT TCT GAG    2213
Trp Leu Ala Val Ser Pro Pro Phe Val Asp Ile Asp Glu His Ser Glu
            635                 640                 645

CAT GGC CAC ATC ATG ATT GTG TGC AAC AAG GGC TCC ATT ATG GCA TTC    2261
His Gly His Ile Met Ile Val Cys Asn Lys Gly Ser Ile Met Ala Phe
        650                 655                 660

TAC TGT GTC CTA GGA TAC TTG GCC TGC CTG GCG CTT GGA AGC TTC ACT    2309
Tyr Cys Val Leu Gly Tyr Leu Ala Cys Leu Ala Leu Gly Ser Phe Thr
    665                 670                 675

ACA GCT TTC TTG GCA AAG AAT CTG CCA GAC ACA TTC AAC GAA GCC AAG    2357
Thr Ala Phe Leu Ala Lys Asn Leu Pro Asp Thr Phe Asn Glu Ala Lys
```

```
680                 685                 690                 695

TTC TTG ACC TTC AGC ATG CTA GTG TTC TGC AGT GTC TGG GTC ACC TTT    2405
Phe Leu Thr Phe Ser Met Leu Val Phe Cys Ser Val Trp Val Thr Phe
                700                 705                 710

CTC CCT GTG TAC CAT AGC ACA AGG GGC AGG GTC ATG GTT GCT GTT GAG    2453
Leu Pro Val Tyr His Ser Thr Arg Gly Arg Val Met Val Ala Val Glu
            715                 720                 725

ATC TTC TCT ATC TTG GCA TCC AGT GCA GGG ATG TTT GGA TGC ATC TTT    2501
Ile Phe Ser Ile Leu Ala Ser Ser Ala Gly Met Phe Gly Cys Ile Phe
        730                 735                 740

GCA CCC AAA ATC TAC ATC ATA TTA ATG AAA CCA GAA AGA AAT TCT ATA    2549
Ala Pro Lys Ile Tyr Ile Ile Leu Met Lys Pro Glu Arg Asn Ser Ile
    745                 750                 755

CAA AAG TTC AGG GAG AAA TCA TAT TTC TAAACAAATA TTTCAGGAAT TTAGTTG  2603
Gln Lys Phe Arg Glu Lys Ser Tyr Phe
760                 765

AATATTAAGT TGGTATATAC CCACCAAATA TTTGGTTATT GTGCATGTAT AGAGTTTTAG   2663
AATCAGTCTT ACTGATTCCT CTATTGCTGT CTAGAGGTAT CTTATCTACC AGTCTTGCAT   2723
ACATTGTCCA TAAAATCTTG TACTCATTCA CTTCTTTAGT TTCCTCTGAG AAAACTAAAT   2783
TTCTCAAATT ATTACTAAAA TGTAATTCAA CATTATGCTT TCATGGATAT TTCCCCCTGG   2843
TTACATCAGA TAAATTTGAT AAGACAGCTG ATTTTGTTAC CTTATATAGA AGGTATATGA   2903
ATGTCCTGCC TTACAGGACA GAGAGGAATT ACACTTAGAA ACCGTCTATC AAGTCAAACA   2963
TTCAATCATA CTGAAAAATA AACTAAAGGA TCAACAGAGA TAAAAAGCAG AATACATTTT   3023
CTGTTTTCTA GTCGGAGCAT ATACATGACA GAATTCTGTT TTTATTTACA GTTGCTCTTC   3083
AAGGTTTTGG TCAATAGTCT AAGATGCAAA TGTTTTCTTT TTTTCTGATC TCAAAAAAAA   3143
TATTATAGCC AACAATTGAA AGAAGCCAGT GACCACTGTG TTTAAATTAG GAACTAGTTT   3203
GAGGATCCTG AGAAGGAGGG TGACTCATTG GAAGACCAGC AGTCTTATCT AACCTGAATA   3263
ACAAAGAATT TTCAGACACT GAGCCTCTAA CCGGGCAGCA TACACCAGTT GATATGAAGC   3323
CCCCAACATA TATGCAACAT AGGATGTCCT GGTCTGGCCT TGGTGAGAGA AGACACACCT   3383
AACCCCCAAG AGACATGATG CTCAAGGGAT TGGGAAGGTG TGGGAGTTGG GAAGGTGGGG   3443
ACTACTTCTT GATGCTGGGA AAGGAGATAT GGGGTGAGGA AGTGTCAGTG CTCAGACTGG   3503
GAAAGGGATA ATGAGTTCAC AGTAAAAAAA ATGTTAAAGA ATAAAAATCT AAAACAAAAT   3563
TAAAAAAAAA AAAAAAAAAA A                                             3584
```

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 768 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```
Met Glu Glu Ile Asn Arg Asn Pro Asp Leu Leu Pro Asn Met Ser Leu
 1               5                  10                  15
Val Ile Lys His Thr Leu Ser Tyr Cys Asp Gly Asn Thr Ala Asp His
            20                  25                  30
Ile Phe Lys Glu Lys Phe Tyr Lys Pro Leu Pro Asn Tyr Val Cys Asn
        35                  40                  45
Glu Glu Thr Met Cys Ser Phe Met Leu Ile Gly Leu Asn Trp Val Leu
    50                  55                  60
Ser Leu Thr Leu Phe Lys Asp Leu Asp Ile Phe Ser Phe Pro Arg Phe
65                  70                  75                  80
Leu Gln Ile Ser Tyr Gly Pro Phe His Ser Ile Phe Ser Asp Asn Glu
                85                  90                  95
```

```
Gln Phe Pro Tyr Leu Tyr Gln Met Thr Pro Lys Asp Thr Ser Leu Ala
                100                 105                 110
Leu Ala Ile Val Ser Phe Leu Leu Tyr Phe Asn Trp Asn Trp Val Gly
            115                 120                 125
Leu Val Ile Ser Asp Asn Asp Glu Gly Asn Gln Phe Leu Ser Glu Leu
        130                 135                 140
Lys Lys Glu Thr Gln Asn Lys Glu Ile Cys Phe Ala Phe Val Asn Met
145                 150                 155                 160
Met Ser Ile His Glu His Ser Ser Tyr Gln Lys Thr Glu Met Tyr Tyr
                165                 170                 175
Asn Gln Ile Val Met Ser Ser Thr Asn Ile Ile Ile Ile Tyr Gly Lys
            180                 185                 190
Thr Asn Ser Ile Ile Glu Leu Ser Phe Arg Met Trp Val Ser Pro Val
        195                 200                 205
Ile Gln Arg Ile Trp Val Thr Asn Ser Glu Leu Asp Phe Pro Thr Ser
    210                 215                 220
Met Arg Asp Phe Thr His Gly Thr Phe Tyr Gly Thr Leu Thr Phe Leu
225                 230                 235                 240
His His His Gly Glu Ile Ser Gly Phe Thr Asn Phe Phe Glu Thr Trp
                245                 250                 255
Asp His Leu Arg Ser Arg Asp Leu Asn Leu Leu Ile Pro Glu Trp Lys
            260                 265                 270
Tyr Phe Ser Tyr Asp Ala Ser Gly Ser Asn Cys Lys Ile Leu Arg Asn
        275                 280                 285
Tyr Ser Ser Asn Ala Ser Leu Glu Trp Ile Thr Glu Gln Lys Phe His
    290                 295                 300
Met Ala Phe Asn Asp Tyr Ser His Ser Ile Tyr Asn Ala Val Tyr Ala
305                 310                 315                 320
Met Ala His Ala Leu His Glu Thr Asn Leu Gln Glu Val Asp Asn Lys
                325                 330                 335
Glu Ile Arg Asn Gly Lys Gly Ala Ser Thr His Cys Leu Lys Val Asn
            340                 345                 350
Ser Phe Leu Arg Lys Thr His Phe Thr Asn Ser His Gly Glu Arg Val
        355                 360                 365
Ile Met Lys Gln Arg Val Arg Val Gln Glu Asp Tyr Asp Ile Val His
    370                 375                 380
Ile Gln Asn Phe Ser Gln His Leu Arg Ile Lys Met Lys Ile Gly Lys
385                 390                 395                 400
Phe Ser Pro Tyr Phe Thr His Gly Gly Pro Phe His Leu Tyr Glu Asp
                405                 410                 415
Met Ile Gln Leu Ala Thr Gly Ser Arg Lys Met Pro Ser Ser Val Cys
            420                 425                 430
Ser Ala Asp Cys Ser Pro Gly Phe Arg Lys Ser Trp Lys Glu Gly Met
        435                 440                 445
Ala Pro Cys Cys Phe Ile Cys Ser Leu Cys Pro Glu Asn Glu Ile Ser
    450                 455                 460
Asn Glu Thr Asn Met Asp Gln Cys Val Asn Cys Pro Glu Tyr Gln Tyr
465                 470                 475                 480
Ala Asn Thr Glu Lys Asn Lys Cys Ile Gln Lys Asp Val Ile Phe Leu
                485                 490                 495
Ser Tyr Glu Asp Pro Leu Gly Met Ala Leu Ala Leu Ile Ala Phe Cys
            500                 505                 510
Leu Ser Ala Phe Thr Ala Val Val Leu Trp Val Phe Val Lys His His
        515                 520                 525
Asp Thr Pro Ile Val Lys Ala Asn Asn Arg Ile Leu Ser Tyr Ile Leu
    530                 535                 540
Ile Met Ser Leu Met Phe Cys Phe Leu Cys Ser Phe Phe Phe Ile Gly
545                 550                 555                 560
His Pro Asn Arg Gly Thr Cys Ile Leu Gln Gln Ile Thr Phe Gly Ile
                565                 570                 575
Val Phe Thr Val Ala Val Ser Thr Val Leu Ala Lys Thr Ile Thr Val
            580                 585                 590
Ile Leu Ala Phe Lys Leu Arg Asp Pro Gly Arg Ser Leu Arg Asn Phe
        595                 600                 605
Leu Val Ser Gly Ala Pro Asn Tyr Ile Ile Pro Ile Cys Ser Leu Leu
```

```
    610                 615                 620
Gln Cys Ile Leu Cys Ala Ile Trp Leu Ala Val Ser Pro Pro Phe Val
625                 630                 635                 640
Asp Ile Asp Glu His Ser Glu His Gly His Ile Met Ile Val Cys Asn
                645                 650                 655
Lys Gly Ser Ile Met Ala Phe Tyr Cys Val Leu Gly Tyr Leu Ala Cys
            660                 665                 670
Leu Ala Leu Gly Ser Phe Thr Thr Ala Phe Leu Ala Lys Asn Leu Pro
        675                 680                 685
Asp Thr Phe Asn Glu Ala Lys Phe Leu Thr Phe Ser Met Leu Val Phe
    690                 695                 700
Cys Ser Val Trp Val Thr Phe Leu Pro Val Tyr His Ser Thr Arg Gly
705                 710                 715                 720
Arg Val Met Val Ala Val Glu Ile Phe Ser Ile Leu Ala Ser Ser Ala
                725                 730                 735
Gly Met Phe Gly Cys Ile Phe Ala Pro Lys Ile Tyr Ile Ile Leu Met
            740                 745                 750
Lys Pro Glu Arg Asn Ser Ile Gln Lys Phe Arg Glu Lys Ser Tyr Phe
        755                 760                 765
```

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3578 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1181...3181
(D) OTHER INFORMATION: GoVN3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

```
CTATCTTGAA GAGTGCTTTT CTGTGTAACT TGCTTTGCTG CACGTTTACA AATTATTTTT     60
TCTTGGTGAA ATTACTAAGA TGTTCTCTTT TCTGTTTGCA ATTCTTGTCC TGAAGCTTTC    120
TTTTCCTTTG TGCAGTCCAA TTGACAACCG TTGTTTTTGG AGATTAAAAA CCAAGACATT    180
TTGGGAAGGA GACAAAGAAC TTGATTGCTT TTTTTTTATT TATACAAGGT TTGGTCATGT    240
AAAGAATGAA CAGTTCAGTG GGAATCTAGA CAAGCGGTTG ACATCTAAGA CTATCCACTT    300
GATTTTGACT CTTTATTTTG CCCTTGAAGA AATAAACAGG AACCCCCATA TTCTACCTAA    360
CATTTCACTG CTAGTTAAAA TTGAATGTGG GCTGCTAGAT GATTGGACAA TAAACAGTTT    420
ATCTTCTAAA AGAGAAAAAT ATCTTCCTAA CTACTACTGT ATAAATCAGA GAAGATATTT    480
AATTGTACTT ACAGGACCAA TGTGGTTAGC ATCTGTCATA GTTGGGCCAC TCCTATACAT    540
AACTAAGAGG CCAGAGATGG ATCAACTCAA CTCTTCTGGC TCAAATTCTT CCCTAAAGTC    600
ACTAATTGGA TATGGCTTTA CTCAGCTTCT CATTGATTTG CTTTGCTTGA ACAATCACTG    660
CCCATTTGTT TTAGTCTTCT GTCTCCTTTA TATTCTGGCT ACAACTGCCT CTACTGATGC    720
ACATTGAACT GCATGAACTC ACAAATTAAC TCAACACCAT TGCACTGCAT TCTTTGCACT    780
GAGTCTCAAA AGTCTGGTTT AACTCTTCTG CATTGAACTC AACTGACTAA TTAGAACTCA    840
GAAATCTGCA TCCCTCTGTC TCCTGAGTAC TTTGATTAAA GGTGTGTACT ATCACACCTG    900
CACCTAAACT TTTCTATACT AAAAATTTGC TTTATACTAG GCTGACCTTG AACTAAGTGA    960
TCTGCTTGCC TCTGTCTCCT GCCTTCCAAG GAATGCCTAT TTCCCAGCAG GATATTTTTT   1020
GCCTACAAGT CTTCAGATGT GATCCATTAA GTATAGTCAT GTTGCTGGAT TAAAATTCCT   1080
CTACAGATTT AATTTTCTGA TCCTGAGGCT AGTGAAACTT TACTATGGGC CATTTCACCC   1140
TCTCTTGAGC AACCAAGAAC TGTATCCATA TCTTTACCAA  ATG GCT CCT AAG GAC    1195
                                            Met Ala Pro Lys Asp
                                             1               5

ACA TCT CTG GCA CTG GCC ATG GTT TCT TTG TTT GTC CAT TTC AGC TGG     1243
Thr Ser Leu Ala Leu Ala Met Val Ser Leu Phe Val His Phe Ser Trp
                10                  15                  20
```

```
AAC TGG GTA GGA GCT GTT GTT TCA GAT GAT GAC CCA GGT TAT GAA TTT     1291
Asn Trp Val Gly Ala Val Val Ser Asp Asp Asp Pro Gly Tyr Glu Phe
            25                  30                  35

ATC TTG GAA TTG AGA AGA GAA ATG CAA AGG AAC AAT TTT TGT TTA GCA     1339
Ile Leu Glu Leu Arg Arg Glu Met Gln Arg Asn Asn Phe Cys Leu Ala
        40                  45                  50

TTT GTG AGT ATC ATT GTT AGT GAT GAC AAT TTA TTT CTG AAA AGG TAT     1387
Phe Val Ser Ile Ile Val Ser Asp Asp Asn Leu Phe Leu Lys Arg Tyr
    55                  60                  65

AAT ATC TAT TAC AAC CAG ATC AAG ATG TCA TCA GCA AAA GTT GTT ATC     1435
Asn Ile Tyr Tyr Asn Gln Ile Lys Met Ser Ser Ala Lys Val Val Ile
70                  75                  80                  85

ATT TAT GGA GAC AAA GAC TCT CCT CTA CAG GTG AAC TTT AGA CTA TGG     1483
Ile Tyr Gly Asp Lys Asp Ser Pro Leu Gln Val Asn Phe Arg Leu Trp
                90                  95                  100

AAT TTA TTT GAT ATC CAA AGA ATC TGG GTC ACT ACT TCA CAG TGG GAT     1531
Asn Leu Phe Asp Ile Gln Arg Ile Trp Val Thr Thr Ser Gln Trp Asp
            105                 110                 115

ATG ATC ATA AAT AAT GGA AAA TTC CTC CTT AAT TCC TTC TAT GGG ACT     1579
Met Ile Ile Asn Asn Gly Lys Phe Leu Leu Asn Ser Phe Tyr Gly Thr
        120                 125                 130

CTC AGT TTT TCA CAT CAC TAT TCT GAA TTA TCT GGT TTT AAA ACA TTT     1627
Leu Ser Phe Ser His His Tyr Ser Glu Leu Ser Gly Phe Lys Thr Phe
    135                 140                 145

ATC CAG ACA GCA TAC CCT TCA AAC TAC AGT GAT GAC TTT TCT CTT GGT     1675
Ile Gln Thr Ala Tyr Pro Ser Asn Tyr Ser Asp Asp Phe Ser Leu Gly
150                 155                 160                 165

ATA TTA TGG TGG GTG TAT TTT AAT TGT TCT TTG TCA TTA TCT GAA TGT     1723
Ile Leu Trp Trp Val Tyr Phe Asn Cys Ser Leu Ser Leu Ser Glu Cys
                170                 175                 180

AAG AAT CTG CAA AAT TGT CCA AAG GAA AAC ATA TTT AGA TGG TTA TAC     1771
Lys Asn Leu Gln Asn Cys Pro Lys Glu Asn Ile Phe Arg Trp Leu Tyr
            185                 190                 195

AGG CAC CAT TTT GAA ATG TCT TTG AGT GAT ACT ACT TAT GAC CTA TAT     1819
Arg His His Phe Glu Met Ser Leu Ser Asp Thr Thr Tyr Asp Leu Tyr
        200                 205                 210

AAT TCT ATG TAT GCT GTG GCT TAC ACA CTC CAA CAG ATG CTT CTG AAA     1867
Asn Ser Met Tyr Ala Val Ala Tyr Thr Leu Gln Gln Met Leu Leu Lys
    215                 220                 225

CAA GCA GAT ACA TGG CAA ATA GAT GAT GGA AAA GAA CCA GAA TTT GAC     1915
Gln Ala Asp Thr Trp Gln Ile Asp Asp Gly Lys Glu Pro Glu Phe Asp
230                 235                 240                 245

TCT TGG CAG ATG CTC TCT TTC CTG AGA AAT ATC CAA TTT ATA AAC CCT     1963
Ser Trp Gln Met Leu Ser Phe Leu Arg Asn Ile Gln Phe Ile Asn Pro
                250                 255                 260

GTT GGT GAC AAA GTG AAC CTG AAT CAT GAA GAA AAA CTG GAT ACA AAG     2011
Val Gly Asp Lys Val Asn Leu Asn His Glu Glu Lys Leu Asp Thr Lys
            265                 270                 275

TAT GAG ATT CAC CAG ACT TTG ACT TTT TTG CCA AAT CCT GTA TTT AAG     2059
```

```
Tyr Glu Ile His Gln Thr Leu Thr Phe Leu Pro Asn Pro Val Phe Lys
        280                 285                 290

CTG AAA ATA GGA ACA TTT TCC CAA AAC TTA TCA CAT GGT CGA CAA TTA     2107
Leu Lys Ile Gly Thr Phe Ser Gln Asn Leu Ser His Gly Arg Gln Leu
    295                 300                 305

TAT ATG TTG AAA GAA ATG ATA GAG TGG AAC ACA GGC CAC CAA CAG TCT     2155
Tyr Met Leu Lys Glu Met Ile Glu Trp Asn Thr Gly His Gln Gln Ser
310                 315                 320                 325

CCA ACC TCA GTT TGC AGT ATT CCT TGT AGT CCA GGA TTC AGA AAA TCC     2203
Pro Thr Ser Val Cys Ser Ile Pro Cys Ser Pro Gly Phe Arg Lys Ser
                330                 335                 340

CCT CAG CTG GGA AAG CCT GTT TGC TGT TTT GAT TGT ACA CCC TGC CCA     2251
Pro Gln Leu Gly Lys Pro Val Cys Cys Phe Asp Cys Thr Pro Cys Pro
            345                 350                 355

GAA AAT GAA ATT TCC AAC ATG ACA AAC ATG AAT CAA TGT ATC AAG TGT     2299
Glu Asn Glu Ile Ser Asn Met Thr Asn Met Asn Gln Cys Ile Lys Cys
        360                 365                 370

CTA AAT GAT CAG TAT GCC AAT CCT GGA GGA ACT CGC TGC CTC AAA AAA     2347
Leu Asn Asp Gln Tyr Ala Asn Pro Gly Gly Thr Arg Cys Leu Lys Lys
    375                 380                 385

GTT ATT GTA TTC CTG GGT TAT GAA GAT CCA TTG GGA ATG TCT CTG GCT     2395
Val Ile Val Phe Leu Gly Tyr Glu Asp Pro Leu Gly Met Ser Leu Ala
390                 395                 400                 405

ATC TTG GCT CTG TGC TTC TCT GCT CTC ACA GCT TTT GTA CTT AGT ATC     2443
Ile Leu Ala Leu Cys Phe Ser Ala Leu Thr Ala Phe Val Leu Ser Ile
                410                 415                 420

TTT TTG AAG CAC CAA GAA ACA CCC ACT GTC AAG GCC AAT AAT AGA ACT     2491
Phe Leu Lys His Gln Glu Thr Pro Thr Val Lys Ala Asn Asn Arg Thr
            425                 430                 435

CTC AGC TAT GTT CTA CTC ATC TCC CTC ATC TCT TGT TTT CTC TGC TCC     2539
Leu Ser Tyr Val Leu Leu Ile Ser Leu Ile Ser Cys Phe Leu Cys Ser
        440                 445                 450

TTG CTC TTC ATT GGT CAT CCC AGC TTT ACC ACA TGT ATC ATG CAG CAG     2587
Leu Leu Phe Ile Gly His Pro Ser Phe Thr Thr Cys Ile Met Gln Gln
    455                 460                 465

ACC ACA TTT GCT GTT GTG TTC ACT GTA GCT GCA TCT ACT GTC TTG GCC     2635
Thr Thr Phe Ala Val Val Phe Thr Val Ala Ala Ser Thr Val Leu Ala
470                 475                 480                 485

AAA ACA ATT ATT GTA ATA TTG GCC TTC AAG GTT ACT AAT ACA AGT AGA     2683
Lys Thr Ile Ile Val Ile Leu Ala Phe Lys Val Thr Asn Thr Ser Arg
                490                 495                 500

AAA ATG AGG TGG CTG CTG GTA TCA GGG GCA CCT AAA TTC ATC ATT CCA     2731
Lys Met Arg Trp Leu Leu Val Ser Gly Ala Pro Lys Phe Ile Ile Pro
            505                 510                 515

ATT TGC ACA ATG ATT CAA CTG ATT CTC TGT GGA ATT TGG CTG GGT ACT     2779
Ile Cys Thr Met Ile Gln Leu Ile Leu Cys Gly Ile Trp Leu Gly Thr
        520                 525                 530

TCT CCT CCA TTT GTT GAT GCT GAT GGA CAT GTT GAA AAA GGC CAC ATT     2827
Ser Pro Pro Phe Val Asp Ala Asp Gly His Val Glu Lys Gly His Ile
```

```
    535                 540                 545

TTG ATT TTC TGT AAC AAA GGT TCA ATT CTT GCT TTC TAT TGT GTC CTG    2875
Leu Ile Phe Cys Asn Lys Gly Ser Ile Leu Ala Phe Tyr Cys Val Leu
550                 555                 560                 565

GGA TAC TTA GTC TCC ATT GCC ATT GCA AGT TTC ACC CTT GCA TTC TTC    2923
Gly Tyr Leu Val Ser Ile Ala Ile Ala Ser Phe Thr Leu Ala Phe Phe
                570                 575                 580

GCC AGA AAT CTG CCC GAC ACA TTC AAT GAA GCC AAG TTC CTA ACA TTC    2971
Ala Arg Asn Leu Pro Asp Thr Phe Asn Glu Ala Lys Phe Leu Thr Phe
            585                 590                 595

AGT ATG CTA GTA TTT TGC AGT GTC TGG GTC ACC TTT CTT CCT GTC TAT    3019
Ser Met Leu Val Phe Cys Ser Val Trp Val Thr Phe Leu Pro Val Tyr
        600                 605                 610

CAT AGC ACC AAG GGC AAG TCT ATG GTG GCT GTG GAA GTT TTC TGT ATA    3067
His Ser Thr Lys Gly Lys Ser Met Val Ala Val Glu Val Phe Cys Ile
    615                 620                 625

TTG GCC TCT AGT GCA GGG CTG CTT TTT TGC ATC TTT GCA CCA AAG TGC    3115
Leu Ala Ser Ser Ala Gly Leu Leu Phe Cys Ile Phe Ala Pro Lys Cys
630                 635                 640                 645

TTC ATT ATT TTG TTA AGA CCT GAG AAA AAA TCT TTT CAG AAG TTT CAG    3163
Phe Ile Ile Leu Leu Arg Pro Glu Lys Lys Ser Phe Gln Lys Phe Gln
                650                 655                 660

AAT ATA CAT TCT AAA ATT TAAAACATTC ATTAAATTTT TCTGACACAC TTGCTAGA  3219
Asn Ile His Ser Lys Ile
            665

CCAAACTTAT TCAGAAGACT CCACTGACAC TACTAGTTGA AATCAAATTT TAGATCCAAA   3279
CATGGAATTT GTTCCCAATA AAGAAAGGAA GCACTATGTA TTAGAATTTA AAAACACGTC   3339
TTAAATCTTG GTTCTCATAA ATCAAACTGT ATGATCAGTC ATTTCAATAA CTGTTTGCTG   3399
TATTTCTTAA TTTTATGCTT ATACTTGAAG AATGTAAAGA CTGGGAATTG GTTCTGAGTT   3459
TTATGAATTA ATTTCTAATT TTACTTTCCT TGGAAAAAAT GTCTAGTGTG TGTTGTTGTG   3519
CTCTATAATA AATAATTATG AGATAAATGC AAAAAAAAAA AAAAAAAAAA AAAAAAAAA    3578
```

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 667 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

```
Met Ala Pro Lys Asp Thr Ser Leu Ala Leu Ala Met Val Ser Leu Phe
 1               5                  10                  15
Val His Phe Ser Trp Asn Trp Val Gly Ala Val Val Ser Asp Asp Asp
            20                  25                  30
Pro Gly Tyr Glu Phe Ile Leu Glu Leu Arg Arg Glu Met Gln Arg Asn
        35                  40                  45
Asn Phe Cys Leu Ala Phe Val Ser Ile Ile Val Ser Asp Asp Asn Leu
    50                  55                  60
Phe Leu Lys Arg Tyr Asn Ile Tyr Tyr Asn Gln Ile Lys Met Ser Ser
65                  70                  75                  80
Ala Lys Val Val Ile Ile Tyr Gly Asp Lys Asp Ser Pro Leu Gln Val
```

```
                    85                  90                  95
Asn Phe Arg Leu Trp Asn Leu Phe Asp Ile Gln Arg Ile Trp Val Thr
            100                 105                 110
Thr Ser Gln Trp Asp Met Ile Ile Asn Asn Gly Lys Phe Leu Leu Asn
        115                 120                 125
Ser Phe Tyr Gly Thr Leu Ser Phe Ser His His Tyr Ser Glu Leu Ser
    130                 135                 140
Gly Phe Lys Thr Phe Ile Gln Thr Ala Tyr Pro Ser Asn Tyr Ser Asp
145                 150                 155                 160
Asp Phe Ser Leu Gly Ile Leu Trp Trp Val Tyr Phe Asn Cys Ser Leu
                165                 170                 175
Ser Leu Ser Glu Cys Lys Asn Leu Gln Asn Cys Pro Lys Glu Asn Ile
            180                 185                 190
Phe Arg Trp Leu Tyr Arg His His Phe Glu Met Ser Leu Ser Asp Thr
        195                 200                 205
Thr Tyr Asp Leu Tyr Asn Ser Met Tyr Ala Val Ala Tyr Thr Leu Gln
    210                 215                 220
Gln Met Leu Leu Lys Gln Ala Asp Thr Trp Gln Ile Asp Asp Gly Lys
225                 230                 235                 240
Glu Pro Glu Phe Asp Ser Trp Gln Met Leu Ser Phe Leu Arg Asn Ile
                245                 250                 255
Gln Phe Ile Asn Pro Val Gly Asp Lys Val Asn Leu Asn His Glu Glu
            260                 265                 270
Lys Leu Asp Thr Lys Tyr Glu Ile His Gln Thr Leu Thr Phe Leu Pro
        275                 280                 285
Asn Pro Val Phe Lys Leu Lys Ile Gly Thr Phe Ser Gln Asn Leu Ser
    290                 295                 300
His Gly Arg Gln Leu Tyr Met Leu Lys Glu Met Ile Glu Trp Asn Thr
305                 310                 315                 320
Gly His Gln Gln Ser Pro Thr Ser Val Cys Ser Ile Pro Cys Ser Pro
                325                 330                 335
Gly Phe Arg Lys Ser Pro Gln Leu Gly Lys Pro Val Cys Cys Phe Asp
            340                 345                 350
Cys Thr Pro Cys Pro Glu Asn Glu Ile Ser Asn Met Thr Asn Met Asn
        355                 360                 365
Gln Cys Ile Lys Cys Leu Asn Asp Gln Tyr Ala Asn Pro Gly Gly Thr
    370                 375                 380
Arg Cys Leu Lys Lys Val Ile Val Phe Leu Gly Tyr Glu Asp Pro Leu
385                 390                 395                 400
Gly Met Ser Leu Ala Ile Leu Ala Leu Cys Phe Ser Ala Leu Thr Ala
                405                 410                 415
Phe Val Leu Ser Ile Phe Leu Lys His Gln Glu Thr Pro Thr Val Lys
            420                 425                 430
Ala Asn Asn Arg Thr Leu Ser Tyr Val Leu Leu Ile Ser Leu Ile Ser
        435                 440                 445
Cys Phe Leu Cys Ser Leu Leu Phe Ile Gly His Pro Ser Phe Thr Thr
    450                 455                 460
Cys Ile Met Gln Gln Thr Thr Phe Ala Val Val Phe Thr Val Ala Ala
465                 470                 475                 480
Ser Thr Val Leu Ala Lys Thr Ile Ile Val Ile Leu Ala Phe Lys Val
                485                 490                 495
Thr Asn Thr Ser Arg Lys Met Arg Trp Leu Leu Val Ser Gly Ala Pro
            500                 505                 510
Lys Phe Ile Ile Pro Ile Cys Thr Met Ile Gln Leu Ile Leu Cys Gly
        515                 520                 525
Ile Trp Leu Gly Thr Ser Pro Pro Phe Val Asp Ala Asp Gly His Val
    530                 535                 540
Glu Lys Gly His Ile Leu Ile Phe Cys Asn Lys Gly Ser Ile Leu Ala
545                 550                 555                 560
Phe Tyr Cys Val Leu Gly Tyr Leu Val Ser Ile Ala Ile Ala Ser Phe
                565                 570                 575
Thr Leu Ala Phe Phe Ala Arg Asn Leu Pro Asp Thr Phe Asn Glu Ala
            580                 585                 590
Lys Phe Leu Thr Phe Ser Met Leu Val Phe Cys Ser Val Trp Val Thr
        595                 600                 605
```

```
Phe Leu Pro Val Tyr His Ser Thr Lys Gly Lys Ser Met Val Ala Val
    610                 615                 620
Glu Val Phe Cys Ile Leu Ala Ser Ser Ala Gly Leu Leu Phe Cys Ile
625                 630                 635                 640
Phe Ala Pro Lys Cys Phe Ile Ile Leu Leu Arg Pro Glu Lys Lys Ser
                645                 650                 655
Phe Gln Lys Phe Gln Asn Ile His Ser Lys Ile
            660                 665
```

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4467 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 126...2723
(D) OTHER INFORMATION: GoVN4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

```
CAGGGATGAG GAAACACCTG TAGAAAAGGA AACCTGAATA CAGGTATAGC ATCTTCTTGG     60
CCAGTGTAGA AGATGGGGAT AATTGCTACC TGTTTGCTGA TCTGTGCAGC AATTAACTAC    120
CAATA ATG TCC AGG CTC AGA GCA GGA AAA AAT ATG CTC ACC TTC ATT TTA    170
      Met Ser Arg Leu Arg Ala Gly Lys Asn Met Leu Thr Phe Ile Leu
       1               5                  10                  15

CTC TTC TTT CTC CTG AAC ATT CCA CTT TTT GTG CCT AGT TTT ATT TAT      218
Leu Phe Phe Leu Leu Asn Ile Pro Leu Phe Val Pro Ser Phe Ile Tyr
                20                  25                  30

CCC AGG TGC TTT TGG AGT ATG AAG AAG AAT GAA TAT CAG GAT AGA AAC      266
Pro Arg Cys Phe Trp Ser Met Lys Lys Asn Glu Tyr Gln Asp Arg Asn
            35                  40                  45

CTG GGA ACA GGT TGT ATG TTC TTT ATT CTA GCA GTG CAA CAG CCT ATG      314
Leu Gly Thr Gly Cys Met Phe Phe Ile Leu Ala Val Gln Gln Pro Met
        50                  55                  60

GAA AAA GAG TAT TTC AGT CAT ATT TCG AAT ATA CAA ACA CCT ACT GAA      362
Glu Lys Glu Tyr Phe Ser His Ile Ser Asn Ile Gln Thr Pro Thr Glu
    65                  70                  75

AAC CAA AAG TAT CCT CTC ACC TTG GCT TTT TCC ATG AAT GAA ATC AAC      410
Asn Gln Lys Tyr Pro Leu Thr Leu Ala Phe Ser Met Asn Glu Ile Asn
80                  85                  90                  95

AAC AAC CCT GAT CTT TTG CCA AAT ATG TCT TTA GCA TTT ACA TTC TCA      458
Asn Asn Pro Asp Leu Leu Pro Asn Met Ser Leu Ala Phe Thr Phe Ser
                100                 105                 110

GAA TAT AGT TGT TAT TTG GAA TCC CAC CAC AAA AGA TTA TTT AAT TTT      506
Glu Tyr Ser Cys Tyr Leu Glu Ser His His Lys Arg Leu Phe Asn Phe
            115                 120                 125

TCT TTA AAA AAT CAT GAA ATT CTC CCT AAT TTT ATC TGT ACA AAA GAC      554
Ser Leu Lys Asn His Glu Ile Leu Pro Asn Phe Ile Cys Thr Lys Asp
        130                 135                 140
```

```
ATC AAG TGT GGA GTG GTA CTT ACC GGA CTT AGT TTG GTA ACA ACT GTG      602
Ile Lys Cys Gly Val Val Leu Thr Gly Leu Ser Leu Val Thr Thr Val
    145                 150                 155

ACA CTT CAT ATA ATC CTA AAC AAT TTC ATA TTT CAG CAG TTC CGT CAG      650
Thr Leu His Ile Ile Leu Asn Asn Phe Ile Phe Gln Gln Phe Arg Gln
160                 165                 170                 175

CTT ACT TAT GGA CAC TTT CAT CCT GCT CTG TGT GAT CAT GAA AAT TTT      698
Leu Thr Tyr Gly His Phe His Pro Ala Leu Cys Asp His Glu Asn Phe
                180                 185                 190

CCT CAT CTA TAT CAG ATG GCC TCT GAT GAT ACA TCT CTA GCC CTT GCT      746
Pro His Leu Tyr Gln Met Ala Ser Asp Asp Thr Ser Leu Ala Leu Ala
            195                 200                 205

CTC GTC TCC TTC ATA ATT CAT TTC AGT TGG AAC TGG ATA GGG TTG GCC      794
Leu Val Ser Phe Ile Ile His Phe Ser Trp Asn Trp Ile Gly Leu Ala
        210                 215                 220

ATC TCA GAC AAT GAT CAA GGC ATA CAT TTT CTC TCT TAT TTG AGA AGA      842
Ile Ser Asp Asn Asp Gln Gly Ile His Phe Leu Ser Tyr Leu Arg Arg
    225                 230                 235

GAG ATG GAA AAA AAT ACA GTC TGC TTT GCC TTT GTC AAC ATT ATT CCA      890
Glu Met Glu Lys Asn Thr Val Cys Phe Ala Phe Val Asn Ile Ile Pro
240                 245                 250                 255

GTC AAT ATG AAT TTA TAC ATG TCA AGA GCT GAA GTG TAT TAC AGC CAA      938
Val Asn Met Asn Leu Tyr Met Ser Arg Ala Glu Val Tyr Tyr Ser Gln
                260                 265                 270

GTT ATG ACA TCA TCC GCA AAT GTT GTT ATC ATT TAT GGT GAT ACA GGG      986
Val Met Thr Ser Ser Ala Asn Val Val Ile Ile Tyr Gly Asp Thr Gly
            275                 280                 285

AAT ACG TTA GCT GTG AGC TTT AGA ATG TGG GAC TCT CTA GGT ATA CAG     1034
Asn Thr Leu Ala Val Ser Phe Arg Met Trp Asp Ser Leu Gly Ile Gln
        290                 295                 300

AGA CTA TGG GTC ACC ACC TCA CAG TGG GAT GTC ACT CCT TTT AAG AAA     1082
Arg Leu Trp Val Thr Thr Ser Gln Trp Asp Val Thr Pro Phe Lys Lys
    305                 310                 315

GAC TTC ACA TTT GAT AAT GGA TAT GGA ACT TTT GGT TTT GGA CAC CGC     1130
Asp Phe Thr Phe Asp Asn Gly Tyr Gly Thr Phe Gly Phe Gly His Arg
320                 325                 330                 335

CAC AGT GAG ATT TCT GGT TTT AAA TAT TTT GTT CAG ACA TTG AAC CCT     1178
His Ser Glu Ile Ser Gly Phe Lys Tyr Phe Val Gln Thr Leu Asn Pro
                340                 345                 350

TTC AAA TAC TCA GAT GAA TAT TTG GTA AAG CTG GAA TGG ATG TAT GTT     1226
Phe Lys Tyr Ser Asp Glu Tyr Leu Val Lys Leu Glu Trp Met Tyr Val
            355                 360                 365

AAT TGT AAA ATC TTA GAA TAT AAC TGT AAG TCA CTG AAG AAC TGC TCC     1274
Asn Cys Lys Ile Leu Glu Tyr Asn Cys Lys Ser Leu Lys Asn Cys Ser
        370                 375                 380

TTT AAT CAC TCA TTG GAA TGG CTA ATG ACA CAT ACT TTT GAC ATG GCC     1322
Phe Asn His Ser Leu Glu Trp Leu Met Thr His Thr Phe Asp Met Ala
    385                 390                 395

ATT ATT GAA GGG AGT TAT GAA ATA TAC AAT GCT GTG TAT GCT TTT GCC     1370
```

```
Ile Ile Glu Gly Ser Tyr Glu Ile Tyr Asn Ala Val Tyr Ala Phe Ala
400                 405                 410                 415

CAT GCA CTC CAT GAG ATG ACT CTT CAA AAT GTT GAT AAT GTT CTC CTT    1418
His Ala Leu His Glu Met Thr Leu Gln Asn Val Asp Asn Val Leu Leu
                420                 425                 430

CCC AAT TAT GAA GAA CAA AAT TAT AAT TGC AAG ATG GTT TAT TCC TTT    1466
Pro Asn Tyr Glu Glu Gln Asn Tyr Asn Cys Lys Met Val Tyr Ser Phe
            435                 440                 445

CTG AGC AAG ACT CAA TTC ACA AAT CCT GTT GGA GAC ACT GTG AAT ATG    1514
Leu Ser Lys Thr Gln Phe Thr Asn Pro Val Gly Asp Thr Val Asn Met
        450                 455                 460

AAT CAA AGA AAC AAA CTG AAG GAA GAG TAC GAC ATT TTC TAC AAT TGG    1562
Asn Gln Arg Asn Lys Leu Lys Glu Glu Tyr Asp Ile Phe Tyr Asn Trp
    465                 470                 475

AAT TTT CCA CAG GGA CTT GGA TTT AAA GTG AAA ATA GGA ATA TTT AGT    1610
Asn Phe Pro Gln Gly Leu Gly Phe Lys Val Lys Ile Gly Ile Phe Ser
480                 485                 490                 495

CCA TAT TTT CCA AAA GGT CAA CAG CTT CAT TTA TCT GAA AAT CTG ATA    1658
Pro Tyr Phe Pro Lys Gly Gln Gln Leu His Leu Ser Glu Asn Leu Ile
                500                 505                 510

GAG TGG TCC ACA GGA CGT ATA CAG ATG CCA ACC TCT GTG TGC AGT GCC    1706
Glu Trp Ser Thr Gly Arg Ile Gln Met Pro Thr Ser Val Cys Ser Ala
            515                 520                 525

GAT TGT GGT CCT GGA TTT AGG AAA GTC TGG AAG AAT GGA ATG CCA GCC    1754
Asp Cys Gly Pro Gly Phe Arg Lys Val Trp Lys Asn Gly Met Pro Ala
        530                 535                 540

TGT TGT TTT GAC TGC AGT CCC TGC CCA GAA AAT GAA ATT TCT AAT GAG    1802
Cys Cys Phe Asp Cys Ser Pro Cys Pro Glu Asn Glu Ile Ser Asn Glu
    545                 550                 555

ACA AAT GTG GAA TTG TGT GTC CAG TGT CCA GAG GAC CAA TAT GCT AAC    1850
Thr Asn Val Glu Leu Cys Val Gln Cys Pro Glu Asp Gln Tyr Ala Asn
560                 565                 570                 575

CAA GAG CAG AAT CAC TGC ATT CAC AAA GCT CGT ATC TTT CTC TCT TAT    1898
Gln Glu Gln Asn His Cys Ile His Lys Ala Arg Ile Phe Leu Ser Tyr
                580                 585                 590

GAT GAA CCC TTG GGG ATG GCT CTT TCC TTA ATG GCC TTA TGC CTC GCT    1946
Asp Glu Pro Leu Gly Met Ala Leu Ser Leu Met Ala Leu Cys Leu Ala
            595                 600                 605

GCA CTC ACA GTT GTG GTT CTT GGA GTC TTT GTG AAA CAT CAC AGA ACT    1994
Ala Leu Thr Val Val Val Leu Gly Val Phe Val Lys His His Arg Thr
        610                 615                 620

CCC ATA GTT AAG GCC AAT AAC TGC ACT CTC ACC TAC ATC TTG CTC ATC    2042
Pro Ile Val Lys Ala Asn Asn Cys Thr Leu Thr Tyr Ile Leu Leu Ile
    625                 630                 635

GCA CTC ATC TTT TGT TTC CTC TGC CCC TTG TTC TTC ATT GGC CAT CCA    2090
Ala Leu Ile Phe Cys Phe Leu Cys Pro Leu Phe Phe Ile Gly His Pro
640                 645                 650                 655

AAC TCA GCT ACC TGC ATC CTT CAG CAA ATC ACA TTT GGA GTT GTG TTC    2138
Asn Ser Ala Thr Cys Ile Leu Gln Gln Ile Thr Phe Gly Val Val Phe
```

```
                660                 665                 670

ACT GTG GCT ATT TCC ACT GTG TTG GCC AAA ACA ACC ACT GTC ATT CTG     2186
Thr Val Ala Ile Ser Thr Val Leu Ala Lys Thr Thr Thr Val Ile Leu
            675                 680                 685

GCT TTC AGA GTC ACA GCC CCT CAT AGA ATG ATG AAG TAC TTT CTT GTT     2234
Ala Phe Arg Val Thr Ala Pro His Arg Met Met Lys Tyr Phe Leu Val
        690                 695                 700

TCA AGG GCA TCT AAC TAC ATC ATT CCC ATT TGT ACT CTC ATT CAA ATT     2282
Ser Arg Ala Ser Asn Tyr Ile Ile Pro Ile Cys Thr Leu Ile Gln Ile
    705                 710                 715

ATT GTA TGT GCC ATC TGG CTA GGA GCT TCT CCT CCT TCT GTT GAT ATT     2330
Ile Val Cys Ala Ile Trp Leu Gly Ala Ser Pro Pro Ser Val Asp Ile
720                 725                 730                 735

GAT GCA CAG TCT GAG CAT GGT CAC ATC ATC ATT GCT TGC AAC AAG GGT     2378
Asp Ala Gln Ser Glu His Gly His Ile Ile Ile Ala Cys Asn Lys Gly
                740                 745                 750

TCA GTC ACT GCT TTT TAC TGT GTC CTG GGA TAT CTG GCC TGC CTG GCC     2426
Ser Val Thr Ala Phe Tyr Cys Val Leu Gly Tyr Leu Ala Cys Leu Ala
            755                 760                 765

TTT GTG AGC TTC ACC CTG GCT TTC CTT TCC AGA AAC CTG CCT GTC ACC     2474
Phe Val Ser Phe Thr Leu Ala Phe Leu Ser Arg Asn Leu Pro Val Thr
        770                 775                 780

TTC AAT GAA GCC AAG TCC ATG ACA TTC AGC ATG CTG GTG TTC TGC AGT     2522
Phe Asn Glu Ala Lys Ser Met Thr Phe Ser Met Leu Val Phe Cys Ser
    785                 790                 795

GTC TGG GTC ACT TTC CTA CCT GTT TAC CAT GGC ACC AAA GGC AAG GTT     2570
Val Trp Val Thr Phe Leu Pro Val Tyr His Gly Thr Lys Gly Lys Val
800                 805                 810                 815

ATG GTG GCT GTT GAG ATC TTT TCC ACC TTG GCT TCT AGT GCA GGA ATG     2618
Met Val Ala Val Glu Ile Phe Ser Thr Leu Ala Ser Ser Ala Gly Met
                820                 825                 830

TTG GGA TGC ATT TTT GCT CCA AAA TGC TAC ACA ATA CTG TTT AGA CCA     2666
Leu Gly Cys Ile Phe Ala Pro Lys Cys Tyr Thr Ile Leu Phe Arg Pro
            835                 840                 845

GAC AGA AAT TCT CTT CAA ATG ATC AGG GAG AAG TCA TCT TCT CAT ACT     2714
Asp Arg Asn Ser Leu Gln Met Ile Arg Glu Lys Ser Ser Ser His Thr
        850                 855                 860

CAC ATT TTA TAAAGTCTGA CTGACACAGG CATTGTTGGT TCATAATCAC CAAATATTC  2772
His Ile Leu
    865

GATTACATTG CCATATCTAT TTTTAGAATG ACTGTCACTG TTCCCTTTGA TGATATTGCG   2832
TAGCAAGATC ATGTCTACTG AGGACTACCT TATCTCCTAT AATCTTCCAA CATTTTCTAC   2892
ATCAATCCTA CTCTTTTAGA GAAAGAGATA ATAGAATTTT AAACATTTTC AGAATTAGAG   2952
TTCTTCTAGG AACAGAGAAG AGAAAGAATT ATTTTTTCAA CAGGTTGATA GAATATCAGG   3012
AAAGGGGTTG AAGTCACAAC AATATAAATA AAGCCCTGCT CTTGTATAGG AACTTATGAA   3072
TACTCAATCC CACCAACTAC CATTAACAAC CACATGTAAC AAATGTTAAA AAGGATCAGA   3132
TGGTTTCTTA TTGTCTCCAA ATTTGCCTGA ACTTATTTAT GCACATAATG AGACACACAC   3192
ACACACACAC ACAAACACAC ACACAAATAC AAATTCCATA AAATTTTAAA AATATAGAAT   3252
ATTACAAAGA CTTAACACTG GCAATCTGCT CTTCAATGTT CATAATTACA GGAACTTACA   3312
GGAAAATATG GGACATAGGT AGAGATGACT GGGTTTATGT TAAGTCATTT TAAATAAGAA   3372
CCCTCAATTT TAAGTGTATC ATAAAAGACA CAGTTGTGAA ATTTTCAAGG ACAGCACTAC   3432
```

```
TTGTTGAAAT AATCTCCATC TGTGGAATTT ATAGGGTTTT GTGACAAAGA TCAGTTCTGA   3492
TATCAGAGAG TAAACTGAAG CAGGCAACCA TTAGTTGTCA GCACTGACAG CAGCTAATGG   3552
AGGTTGCTTC AGAAATCAAT TGAGGTTGAT TCTGGCAATG AGCAGTTAGA GAAGATAAAA   3612
AACAGGGAAA TCAAATATTC ACACACACAC ACACACACAC ACGTACACTC ACATGCACAA   3672
GCAAGTGCAT GCATGCAAAC CCACACAGAC TACTTGAAGC AAAGGCAAGG TCCAGCCACT   3732
TGAAACATAC AAATGTGTAC ATATAGACAG ACACAGACAA ACACATACAT ATCCACATGT   3792
TAAATGGCTG GAGCAATGTC AGCCAGCAGG CTCCATGTAT TTCACATATG TACATATATG   3852
CATGTAAATA AATATTCAGA TATACACATA TTCACATGTA CTGGTGGGTA GGTGGAATAA   3912
AGTTCCAAAA AACAGGCCCC AGGAATTTTA CACATAATGT ACAGACATAT ATAACACTAT   3972
TGGTGGAAGA ACAAGCTCCA ACATATTCAG GGAAGCATTG CATATACATA CATATAGATT   4032
TGATGGATGG AACAAAGTTC CAACAAATTC TCACATGAAC TTTATATATG TATATACATG   4092
AAAGGCAGCC TGGTTCCCAG TTGATCAGAG GTTTGAAAGC CCAGTGACCC TAAAAAAGAT   4152
GGTAGCCATT TAGCCTGATT CCCAGTAAAC CAGGCAAGTC ACTAGCCACA GCCCTCCATA   4212
GAATTTTGGC CATCAGTCAC TTAAGCCCAA CACCCTCCAC AGATTAAAGG AAGTGATTAC   4272
AGGTCACAGG GACTCAGAAC ACATTTCCAT TATGTGACAT AGTCAAAGAC TTGGAGACTT   4332
AGCCAATGAA CTTTCCTTCC CTGAAACTCC TCCCTGCAGG CCAACCTTGA AAAGAGGGGT   4392
ATGGTTTTAC TCATCTGCTT TCAGCCATGA CAATAAATGA CTTAAAACAA TGAAAAAAAA   4452
AAAAAAAAAA AAAAA                                                    4467
```

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 866 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

```
Met Ser Arg Leu Arg Ala Gly Lys Asn Met Leu Thr Phe Ile Leu Leu
 1               5                  10                  15
Phe Phe Leu Leu Asn Ile Pro Leu Phe Val Pro Ser Phe Ile Tyr Pro
            20                  25                  30
Arg Cys Phe Trp Ser Met Lys Lys Asn Glu Tyr Gln Asp Arg Asn Leu
        35                  40                  45
Gly Thr Gly Cys Met Phe Phe Ile Leu Ala Val Gln Gln Pro Met Glu
    50                  55                  60
Lys Glu Tyr Phe Ser His Ile Ser Asn Ile Gln Thr Pro Thr Glu Asn
65                  70                  75                  80
Gln Lys Tyr Pro Leu Thr Leu Ala Phe Ser Met Asn Glu Ile Asn Asn
                85                  90                  95
Asn Pro Asp Leu Leu Pro Asn Met Ser Leu Ala Phe Thr Phe Ser Glu
            100                 105                 110
Tyr Ser Cys Tyr Leu Glu Ser His His Lys Arg Leu Phe Asn Phe Ser
        115                 120                 125
Leu Lys Asn His Glu Ile Leu Pro Asn Phe Ile Cys Thr Lys Asp Ile
    130                 135                 140
Lys Cys Gly Val Val Leu Thr Gly Leu Ser Leu Val Thr Thr Val Thr
145                 150                 155                 160
Leu His Ile Ile Leu Asn Asn Phe Ile Phe Gln Gln Phe Arg Gln Leu
                165                 170                 175
Thr Tyr Gly His Phe His Pro Ala Leu Cys Asp His Glu Asn Phe Pro
            180                 185                 190
His Leu Tyr Gln Met Ala Ser Asp Asp Thr Ser Leu Ala Leu Ala Leu
        195                 200                 205
Val Ser Phe Ile Ile His Phe Ser Trp Asn Trp Ile Gly Leu Ala Ile
    210                 215                 220
Ser Asp Asn Asp Gln Gly Ile His Phe Leu Ser Tyr Leu Arg Arg Glu
225                 230                 235                 240
Met Glu Lys Asn Thr Val Cys Phe Ala Phe Val Asn Ile Ile Pro Val
                245                 250                 255
Asn Met Asn Leu Tyr Met Ser Arg Ala Glu Val Tyr Tyr Ser Gln Val
```

```
                260                 265                 270
Met Thr Ser Ser Ala Asn Val Val Ile Ile Tyr Gly Asp Thr Gly Asn
        275                 280                 285
Thr Leu Ala Val Ser Phe Arg Met Trp Asp Ser Leu Gly Ile Gln Arg
    290                 295                 300
Leu Trp Val Thr Thr Ser Gln Trp Asp Val Thr Pro Phe Lys Lys Asp
305                 310                 315                 320
Phe Thr Phe Asp Asn Gly Tyr Gly Thr Phe Gly Phe Gly His Arg His
                325                 330                 335
Ser Glu Ile Ser Gly Phe Lys Tyr Phe Val Gln Thr Leu Asn Pro Phe
            340                 345                 350
Lys Tyr Ser Asp Glu Tyr Leu Val Lys Leu Glu Trp Met Tyr Val Asn
        355                 360                 365
Cys Lys Ile Leu Glu Tyr Asn Cys Lys Ser Leu Lys Asn Cys Ser Phe
    370                 375                 380
Asn His Ser Leu Glu Trp Leu Met Thr His Thr Phe Asp Met Ala Ile
385                 390                 395                 400
Ile Glu Gly Ser Tyr Glu Ile Tyr Asn Ala Val Tyr Ala Phe Ala His
                405                 410                 415
Ala Leu His Glu Met Thr Leu Gln Asn Val Asp Asn Val Leu Leu Pro
            420                 425                 430
Asn Tyr Glu Glu Gln Asn Tyr Asn Cys Lys Met Val Tyr Ser Phe Leu
        435                 440                 445
Ser Lys Thr Gln Phe Thr Asn Pro Val Gly Asp Thr Val Asn Met Asn
    450                 455                 460
Gln Arg Asn Lys Leu Lys Glu Glu Tyr Asp Ile Phe Tyr Asn Trp Asn
465                 470                 475                 480
Phe Pro Gln Gly Leu Gly Phe Lys Val Lys Ile Gly Ile Phe Ser Pro
                485                 490                 495
Tyr Phe Pro Lys Gly Gln Gln Leu His Leu Ser Glu Asn Leu Ile Glu
            500                 505                 510
Trp Ser Thr Gly Arg Ile Gln Met Pro Thr Ser Val Cys Ser Ala Asp
        515                 520                 525
Cys Gly Pro Gly Phe Arg Lys Val Trp Lys Asn Gly Met Pro Ala Cys
    530                 535                 540
Cys Phe Asp Cys Ser Pro Cys Pro Glu Asn Glu Ile Ser Asn Glu Thr
545                 550                 555                 560
Asn Val Glu Leu Cys Val Gln Cys Pro Glu Asp Gln Tyr Ala Asn Gln
                565                 570                 575
Glu Gln Asn His Cys Ile His Lys Ala Arg Ile Phe Leu Ser Tyr Asp
            580                 585                 590
Glu Pro Leu Gly Met Ala Leu Ser Leu Met Ala Leu Cys Leu Ala Ala
        595                 600                 605
Leu Thr Val Val Val Leu Gly Val Phe Val Lys His His Arg Thr Pro
    610                 615                 620
Ile Val Lys Ala Asn Asn Cys Thr Leu Thr Tyr Ile Leu Leu Ile Ala
625                 630                 635                 640
Leu Ile Phe Cys Phe Leu Cys Pro Leu Phe Phe Ile Gly His Pro Asn
                645                 650                 655
Ser Ala Thr Cys Ile Leu Gln Gln Ile Thr Phe Gly Val Val Phe Thr
            660                 665                 670
Val Ala Ile Ser Thr Val Leu Ala Lys Thr Thr Thr Val Ile Leu Ala
        675                 680                 685
Phe Arg Val Thr Ala Pro His Arg Met Met Lys Tyr Phe Leu Val Ser
    690                 695                 700
Arg Ala Ser Asn Tyr Ile Ile Pro Ile Cys Thr Leu Ile Gln Ile Ile
705                 710                 715                 720
Val Cys Ala Ile Trp Leu Gly Ala Ser Pro Pro Ser Val Asp Ile Asp
                725                 730                 735
Ala Gln Ser Glu His Gly His Ile Ile Ile Ala Cys Asn Lys Gly Ser
            740                 745                 750
Val Thr Ala Phe Tyr Cys Val Leu Gly Tyr Leu Ala Cys Leu Ala Phe
        755                 760                 765
Val Ser Phe Thr Leu Ala Phe Leu Ser Arg Asn Leu Pro Val Thr Phe
    770                 775                 780
```

```
Asn Glu Ala Lys Ser Met Thr Phe Ser Met Leu Val Phe Cys Ser Val
785                 790                 795                 800
Trp Val Thr Phe Leu Pro Val Tyr His Gly Thr Lys Gly Lys Val Met
                805                 810                 815
Val Ala Val Glu Ile Phe Ser Thr Leu Ala Ser Ser Ala Gly Met Leu
            820                 825                 830
Gly Cys Ile Phe Ala Pro Lys Cys Tyr Thr Ile Leu Phe Arg Pro Asp
        835                 840                 845
Arg Asn Ser Leu Gln Met Ile Arg Glu Lys Ser Ser Ser His Thr His
    850                 855                 860
Ile Leu
865
```

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2916 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 299...2635
(D) OTHER INFORMATION: GoVN5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

```
CGGCACGAGT TCAACTAGTC ATGTTCAAGA AGGGGCAAAT ACTTTGTTAA TATGCTCTTC     60
GCTTGGACTT TTATCTCTTG CTTTCTGCAG ATTCCAATTA TTTTATGCTC CTACAGAAGC    120
AGCGAGTGCT TAGTCAAGAT GAATTATCGT TTAAAGGGGA AAGGAAATGT GGTGATTGTT    180
GGATTTTTCC CTGCTTTTGC TGTCTACCCC CTCAACAAAA CAATTGACTG GTGGATGCTT    240
AAATTCAGCA AAGAATTATG ATTGAGTTTA AGTTGAAGAG CTACCAGTAT ATTTGGCC AT   300
                                                                 Met
                                                                  1

GAG GTT TGC CAT TGA GGA AAT CAA CAG CAA TCC CCA TCT TTT ACC AAA      348
 Arg Phe Ala Ile Glu Glu Ile Asn Ser Asn Pro His Leu Leu Pro Asn
           5                  10                  15

CAC ATC CCT GGG ATT TGA GAT CAA TAA TGT CCC ACA CGG TCA GAG GTA      396
 Thr Ser Leu Gly Phe Glu Ile Asn Asn Val Pro His Gly Gln Arg Tyr
       20                  25                  30

CAC TCT GGT CAA ACT TTT TAG CTC ACT TTC AGG GTC TAA TTA TGA CAT      444
 Thr Leu Val Lys Leu Phe Ser Ser Leu Ser Gly Ser Asn Tyr Asp Ile
   35                  40                  45

TCC TAA CTA CAT AAG TGC AAG TGA GAG CAA TTC TGC TGC TGT ACT TAC      492
 Pro Asn Tyr Ile Ser Ala Ser Glu Ser Asn Ser Ala Ala Val Leu Thr
 50                  55                  60                  65

AGG ACC ATC GTG GAC AAT ATC TGA ATG CGT AGG GAC ACT CCT GGA TCT      540
 Gly Pro Ser Trp Thr Ile Ser Glu Cys Val Gly Thr Leu Leu Asp Leu
                 70                  75                  80

TTA CAA ATT TCC ACA GCT TAC TTT TGG GCC TTT TGA TAG TCT CCT GAG      588
 Tyr Lys Phe Pro Gln Leu Thr Phe Gly Pro Phe Asp Ser Leu Leu Ser
             85                  90                  95

TGA ACA AAG ACG GTT TTC TTC TCT GTA CCA AGT GGC CCC CAA AGA TAC      636
 Glu Gln Arg Arg Phe Ser Ser Leu Tyr Gln Val Ala Pro Lys Asp Thr.
```

```
          100                 105                 110

ATT TCT GAC GCC TGG CAT TGT ATC TTT GAT GCT TCA TTT CCA CTG GAA      684
 Phe Leu Thr Pro Gly Ile Val Ser Leu Met Leu His Phe His Trp Asn
     115                 120                 125

CTG GGT GGG GTT ATT CAT CAT AGA TGA TGA CAA AGG TGC CCA GAC ACT      732
 Trp Val Gly Leu Phe Ile Ile Asp Asp Asp Lys Gly Ala Gln Thr Leu
 130                 135                 140                 14

GTC AGA CTT GAG AAA TGA GAT GGA TAA AAA TGG AGT CTG CAC AGC ATT      780
 Ser Asp Leu Arg Asn Glu Met Asp Lys Asn Gly Val Cys Thr Ala Phe
5                150                 155                 160

TGT AGA AAT GAT CCC AGT CAT CAA GGG TTC ATT TTT TAC CAA ATC CTG      828
 Val Glu Met Ile Pro Val Ile Lys Gly Ser Phe Phe Thr Lys Ser Trp
             165                 170                 175

GAA AAA TCA TGT GCA GAT CCT GGA ATC ATC ATC AAA TGT GAT TAT TAT      876
 Lys Asn His Val Gln Ile Leu Glu Ser Ser Ser Asn Val Ile Ile Ile
         180                 185                 190

TTA TGG GGA CTC TGA TTC TCT ATT AAG CTT AAT AGT AAA TAT TAA GCA      924
 Tyr Gly Asp Ser Asp Ser Leu Leu Ser Leu Ile Val Asn Ile Lys Gln
     195                 200                 205

GAA GTT GCT CAC ATG GAA AGT GTG GGT ACT GAT CTC ACA GTG GGA TGT      972
 Lys Leu Leu Thr Trp Lys Val Trp Val Leu Ile Ser Gln Trp Asp Val
 210                 215                 220                 22

TTC TAA ATT TGA TGA TTA TTT CAT GGT AGA CTC ATT GCA TGG AGC TCT     1020
 Ser Lys Phe Asp Asp Tyr Phe Met Val Asp Ser Leu His Gly Ala Leu
5                230                 235                 240

TAT TTT TTC ACA CCA TCG TGA GGA GAT TCC TAA TTT TAC AGA TTT TAT     1068
 Ile Phe Ser His His Arg Glu Glu Ile Pro Asn Phe Thr Asp Phe Met
             245                 250                 255

GCA GAA GTA CAA CCC TTC CAA GTA CCC GGA AGA CAC TTA TCT TCA TGT     1116
 Gln Lys Tyr Asn Pro Ser Lys Tyr Pro Glu Asp Thr Tyr Leu His Val
         260                 265                 270

ATT GTG GCA CAT GTA CTT CAA TTG CTC ATT TGT TAA GAA AGA TTG TAA     1164
 Leu Trp His Met Tyr Phe Asn Cys Ser Phe Val Lys Lys Asp Cys Lys
     275                 280                 285

AAT TGT GCA CAA CTG TTT GCC TAA TGC CTC CCT GGG GTT CTT GCC TGG     1212
 Ile Val His Asn Cys Leu Pro Asn Ala Ser Leu Gly Phe Leu Pro Gly
 290                 295                 300                 30

GAA CAT ATT TGA CAT GGC CAT GAG TGA AGA GAG TTA CAA TGT ATA CAA     1260
 Asn Ile Phe Asp Met Ala Met Ser Glu Glu Ser Tyr Asn Val Tyr Asn
5                310                 315                 320

TGC TGT GTA TGC TGT GGC CCA CAG TCT GCA TGA GAT GAT TCT CAA CCA     1308
 Ala Val Tyr Ala Val Ala His Ser Leu His Glu Met Ile Leu Asn Gln
             325                 330                 335

AGT ACA ATT TCA AAC TCA TGA AAA AGG AAA AAA GAT GGT ATT CTT TCC     1356
 Val Gln Phe Gln Thr His Glu Lys Gly Lys Lys Met Val Phe Phe Pro
         340                 345                 350

TTG GCA GCT TCA CCC CTT TCT AAG GGA AAG ACA ACT CAT CAA TCA GAA     1404
 Trp Gln Leu His Pro Phe Leu Arg Glu Arg Gln Leu Ile Asn Gln Asn
     355                 360                 365
```

```
TGG AGC GAA TGA AGA TCT GGA TTG TAC CAG GAA GTC ACA TGT AGA GTA    1452
 Gly Ala Asn Glu Asp Leu Asp Cys Thr Arg Lys Ser His Val Glu Tyr
 370                 375                 380                 38

TGA CAT TCT CAA CTT TTG GAA TTT CCC AAA AGG TCT TGG GCT AAA TGT    1500
 Asp Ile Leu Asn Phe Trp Asn Phe Pro Lys Gly Leu Gly Leu Asn Val
5                390                 395                 400

GAA AGT AGG AAC GTT TTC TCC AAG TGC TCC AAA GGA ACA GAA ACT GTC    1548
 Lys Val Gly Thr Phe Ser Pro Ser Ala Pro Lys Glu Gln Lys Leu Ser
            405                 410                 415

CAT ATC TTC TAA CAT GAT ACA GTG GGC CAC AGG GTC GAC AGA GAT TCC    1596
 Ile Ser Ser Asn Met Ile Gln Trp Ala Thr Gly Ser Thr Glu Ile Pro
        420                 425                 430

ACA GTC TGT ATG CAG TGA GAG CTG TCA TCC TGG ATT CAG GAA AAC CCA    1644
 Gln Ser Val Cys Ser Glu Ser Cys His Pro Gly Phe Arg Lys Thr His
    435                 440                 445

CCA GGA AGG CAG GGT TGC CTG TTG CTT TGA CTG CAT TCC TTG TCC AGA    1692
 Gln Glu Gly Arg Val Ala Cys Cys Phe Asp Cys Ile Pro Cys Pro Glu
 450                 455                 460                 46

AAA TGA GAT CTC CAA TGA GAC AGA TGT GGA TCA GTG TGT GAA GTG TCC    1740
 Asn Glu Ile Ser Asn Glu Thr Asp Val Asp Gln Cys Val Lys Cys Pro
5                470                 475                 480

AGA AAC TCA CTA TGC AAA CAT AGA GAA GAT CCA CTG CCT ACA GAA AAC    1788
 Glu Thr His Tyr Ala Asn Ile Glu Lys Ile His Cys Leu Gln Lys Thr
            485                 490                 495

TGT GAC ATT TCT GTA CTA TGA TGA CCC ATT GGG GAA GAC ACT TTG CTT    1836
 Val Thr Phe Leu Tyr Tyr Asp Asp Pro Leu Gly Lys Thr Leu Cys Phe
        500                 505                 510

CAT GTC CCT GGG TTT CTC CTC ACT CAC AGC TGC TGT TCT TGT GGT GTT    1884
 Met Ser Leu Gly Phe Ser Ser Leu Thr Ala Ala Val Leu Val Val Phe
    515                 520                 525

TCT GAA GAA CAG GGA CAC CCC CAT TGT CAA GGC CAA TAA CCT GGC TCT    1932
 Leu Lys Asn Arg Asp Thr Pro Ile Val Lys Ala Asn Asn Leu Ala Leu
 530                 535                 540                 54

CAG TTA CAC CCT GCT CAT CAC TTT GAT GCT CTG TTT TCT CTG TCC CTT    1980
 Ser Tyr Thr Leu Leu Ile Thr Leu Met Leu Cys Phe Leu Cys Pro Leu
5                550                 555                 560

GCT CTT CAT TGG CCG TCC CAG CAC AGC CTC CTG TAT CCT GCA GCA AAA    2028
 Leu Phe Ile Gly Arg Pro Ser Thr Ala Ser Cys Ile Leu Gln Gln Asn
            565                 570                 575

CAT TTT TGG GCT TCT GTT CAC TGT GGC TCT TTC CAC TGT GTT GGC CAA    2076
 Ile Phe Gly Leu Leu Phe Thr Val Ala Leu Ser Thr Val Leu Ala Lys
        580                 585                 590

AAC TAT CAC TGT GGT TAT AGC CTT CAA GAT CAC TTC TCC AGG AAG AAT    2124
 Thr Ile Thr Val Val Ile Ala Phe Lys Ile Thr Ser Pro Gly Arg Ile
    595                 600                 605

TAG AAG ATG GCT GCT GAT ATC AAG GGC CCC TAA TTT CAT TAT TCC CTT    2172
 Arg Arg Trp Leu Leu Ile Ser Arg Ala Pro Asn Phe Ile Ile Pro Leu
 610                 615                 620                 62

ATG CAC CCT GCT CCA AGT TTT TCT ATC TGG AAT TTG GCT GAC AAC CTC    2220
```

```
  Cys Thr Leu Leu Gln Val Phe Leu Ser Gly Ile Trp Leu Thr Thr Ser
5                 630                 635                 640

TCC TCC ATT TAT TGA TAA AGA TGC TCA CTC AGA ACA TGG ACA CAT CAT    2268
 Pro Pro Phe Ile Asp Lys Asp Ala His Ser Glu His Gly His Ile Ile
            645                 650                 655

CAT CAT TTG CAA TAA AGG CTC AGC TGT TGC TTT CCA TTG CAA CCT TGG    2316
 Ile Ile Cys Asn Lys Gly Ser Ala Val Ala Phe His Cys Asn Leu Gly
        660                 665                 670

ATA CCT GGG AGC ACT AGC CCT AGT GAG CTA CTT TAT GGC TTT CTT GTC    2364
 Tyr Leu Gly Ala Leu Ala Leu Val Ser Tyr Phe Met Ala Phe Leu Ser
    675                 680                 685

CAG AAA CCT ACC TGA CAC ATT CAA TGA AGC CAA GTT CCT GGC TTT CAG    2412
 Arg Asn Leu Pro Asp Thr Phe Asn Glu Ala Lys Phe Leu Ala Phe Ser
 690                 695                 700                 70

CAT GCT GGT GTT CTG CAG TGT CTG GGT CAC CTT CCT CCC TGT CTA CCA    2460
 Met Leu Val Phe Cys Ser Val Trp Val Thr Phe Leu Pro Val Tyr His
5                 710                 715                 720

CAG CAC CAA GGG GAA GAA CAT GGT GGC TAT GGA AGT CTT CTC TAT CTT    2508
 Ser Thr Lys Gly Lys Asn Met Val Ala Met Glu Val Phe Ser Ile Leu
            725                 730                 735

GGC TTC CAG TAC ATC TCT CCT AGG CAT CAT CTT TGC CCC CAA GTG CTA    2556
 Ala Ser Ser Thr Ser Leu Leu Gly Ile Ile Phe Ala Pro Lys Cys Tyr
        740                 745                 750

CCT CAT ATT ATT AAG ACC AGA AAG GAA TTC ACT TAG CTA TAT CAG GGA    2604
 Leu Ile Leu Leu Arg Pro Glu Arg Asn Ser Leu Ser Tyr Ile Arg Asp
    755                 760                 765

CAA AAC ATA TGC TAA AAG CAT AAA ACC TTC T TAGCATCCTT ATGTGCCTCT T  2656
 Lys Thr Tyr Ala Lys Ser Ile Lys Pro Ser
 770                 775

AAATTAAACA GCATCATTGA AGGCAATTGT TGTTCTTCAC TATCTGAACA CTCACATATA   2716
AAGTCATAAT TGTACATTTG ATCCAGGGGC TATTATTTCT TTAGTAGTCA TATATATGTA   2776
CCTAATGCTT TTTTCACATT AAAATATGTG CTGCATTTTT CGTCTTCCTC TTCTACTTAC   2836
TATTAGTTTT GTGCTATTGA TTTAACTTGC AATAAAATCC AAATTTCTGA GTTCTTCCAA   2896
AAAAAAAAAA AAAAAAAAAA                                               2916
```

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 779 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

```
Met Arg Phe Ala Ile Glu Glu Ile Asn Ser Asn Pro His Leu Leu Pro
 1               5                  10                  15
Asn Thr Ser Leu Gly Phe Glu Ile Asn Asn Val Pro His Gly Gln Arg
            20                  25                  30
Tyr Thr Leu Val Lys Leu Phe Ser Ser Leu Ser Gly Ser Asn Tyr Asp
        35                  40                  45
Ile Pro Asn Tyr Ile Ser Ala Ser Glu Ser Asn Ser Ala Ala Val Leu
```

```
    50                  55                  60
Thr Gly Pro Ser Trp Thr Ile Ser Glu Cys Val Gly Thr Leu Leu Asp
65                  70                  75                  80
Leu Tyr Lys Phe Pro Gln Leu Thr Phe Gly Pro Phe Asp Ser Leu Leu
                85                  90                  95
Ser Glu Gln Arg Arg Phe Ser Ser Leu Tyr Gln Val Ala Pro Lys Asp
            100                 105                 110
Thr Phe Leu Thr Pro Gly Ile Val Ser Leu Met Leu His Phe His Trp
        115                 120                 125
Asn Trp Val Gly Leu Phe Ile Ile Asp Asp Asp Lys Gly Ala Gln Thr
    130                 135                 140
Leu Ser Asp Leu Arg Asn Glu Met Asp Lys Asn Gly Val Cys Thr Ala
145                 150                 155                 160
Phe Val Glu Met Ile Pro Val Ile Lys Gly Ser Phe Phe Thr Lys Ser
                165                 170                 175
Trp Lys Asn His Val Gln Ile Leu Glu Ser Ser Ser Asn Val Ile Ile
            180                 185                 190
Ile Tyr Gly Asp Ser Asp Ser Leu Leu Ser Leu Ile Val Asn Ile Lys
        195                 200                 205
Gln Lys Leu Leu Thr Trp Lys Val Trp Val Leu Ile Ser Gln Trp Asp
    210                 215                 220
Val Ser Lys Phe Asp Asp Tyr Phe Met Val Asp Ser Leu His Gly Ala
225                 230                 235                 240
Leu Ile Phe Ser His His Arg Glu Glu Ile Pro Asn Phe Thr Asp Phe
                245                 250                 255
Met Gln Lys Tyr Asn Pro Ser Lys Tyr Pro Glu Asp Thr Tyr Leu His
            260                 265                 270
Val Leu Trp His Met Tyr Phe Asn Cys Ser Phe Val Lys Lys Asp Cys
        275                 280                 285
Lys Ile Val His Asn Cys Leu Pro Asn Ala Ser Leu Gly Phe Leu Pro
    290                 295                 300
Gly Asn Ile Phe Asp Met Ala Met Ser Glu Glu Ser Tyr Asn Val Tyr
305                 310                 315                 320
Asn Ala Val Tyr Ala Val Ala His Ser Leu His Glu Met Ile Leu Asn
                325                 330                 335
Gln Val Gln Phe Gln Thr His Glu Lys Gly Lys Lys Met Val Phe Phe
            340                 345                 350
Pro Trp Gln Leu His Pro Phe Leu Arg Glu Arg Gln Leu Ile Asn Gln
        355                 360                 365
Asn Gly Ala Asn Glu Asp Leu Asp Cys Thr Arg Lys Ser His Val Glu
    370                 375                 380
Tyr Asp Ile Leu Asn Phe Trp Asn Phe Pro Lys Gly Leu Gly Leu Asn
385                 390                 395                 400
Val Lys Val Gly Thr Phe Ser Pro Ser Ala Pro Lys Glu Gln Lys Leu
                405                 410                 415
Ser Ile Ser Ser Asn Met Ile Gln Trp Ala Thr Gly Ser Thr Glu Ile
            420                 425                 430
Pro Gln Ser Val Cys Ser Glu Ser Cys His Pro Gly Phe Arg Lys Thr
        435                 440                 445
His Gln Glu Gly Arg Val Ala Cys Cys Phe Asp Cys Ile Pro Cys Pro
    450                 455                 460
Glu Asn Glu Ile Ser Asn Glu Thr Asp Val Asp Gln Cys Val Lys Cys
465                 470                 475                 480
Pro Glu Thr His Tyr Ala Asn Ile Glu Lys Ile His Cys Leu Gln Lys
                485                 490                 495
Thr Val Thr Phe Leu Tyr Tyr Asp Asp Pro Leu Gly Lys Thr Leu Cys
            500                 505                 510
Phe Met Ser Leu Gly Phe Ser Ser Leu Thr Ala Ala Val Leu Val Val
        515                 520                 525
Phe Leu Lys Asn Arg Asp Thr Pro Ile Val Lys Ala Asn Asn Leu Ala
    530                 535                 540
Leu Ser Tyr Thr Leu Leu Ile Thr Leu Met Leu Cys Phe Leu Cys Pro
545                 550                 555                 560
Leu Leu Phe Ile Gly Arg Pro Ser Thr Ala Ser Cys Ile Leu Gln Gln
                565                 570                 575
```

```
Asn Ile Phe Gly Leu Leu Phe Thr Val Ala Leu Ser Thr Val Leu Ala
            580                 585                 590
Lys Thr Ile Thr Val Val Ile Ala Phe Lys Ile Thr Ser Pro Gly Arg
        595                 600                 605
Ile Arg Arg Trp Leu Leu Ile Ser Arg Ala Pro Asn Phe Ile Ile Pro
    610                 615                 620
Leu Cys Thr Leu Leu Gln Val Phe Leu Ser Gly Ile Trp Leu Thr Thr
625                 630                 635                 640
Ser Pro Pro Phe Ile Asp Lys Asp Ala His Ser Glu His Gly His Ile
                645                 650                 655
Ile Ile Ile Cys Asn Lys Gly Ser Ala Val Ala Phe His Cys Asn Leu
            660                 665                 670
Gly Tyr Leu Gly Ala Leu Ala Leu Val Ser Tyr Phe Met Ala Phe Leu
        675                 680                 685
Ser Arg Asn Leu Pro Asp Thr Phe Asn Glu Ala Lys Phe Leu Ala Phe
    690                 695                 700
Ser Met Leu Val Phe Cys Ser Val Trp Val Thr Phe Leu Pro Val Tyr
705                 710                 715                 720
His Ser Thr Lys Gly Lys Asn Met Val Ala Met Glu Val Phe Ser Ile
                725                 730                 735
Leu Ala Ser Ser Thr Ser Leu Leu Gly Ile Ile Phe Ala Pro Lys Cys
            740                 745                 750
Tyr Leu Ile Leu Leu Arg Pro Glu Arg Asn Ser Leu Ser Tyr Ile Arg
        755                 760                 765
Asp Lys Thr Tyr Ala Lys Ser Ile Lys Pro Ser
    770                 775
```

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3307 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 112...1761
(D) OTHER INFORMATION: GoVN6

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

```
TAAGGCAGGA AAAAATGTTC ATTTTGATGG AAGTCTTCTT CTTCTTCCTT AACATTCCAC     60
TGCTCATGGC AAATTTCATT GATCCCAAGT GCTTTTGGAG AGTAAATTTG A ATG AAG    117
                                                          Met Lys
                                                           1

TTA AGG GAT AAA GAC TTG AGC ATA ACT TGT TCC TTC ATC CTT GAA GCA     165
Leu Arg Asp Lys Asp Leu Ser Ile Thr Cys Ser Phe Ile Leu Glu Ala
 5                  10                  15

GTT CAG ATG CCT ACG GAA AAC GAT TAT TTC AAC CAG ACT CTG AAT ATC     213
Val Gln Met Pro Thr Glu Asn Asp Tyr Phe Asn Gln Thr Leu Asn Ile
    20                  25                  30

CTA AAA ACA ACA AAA AAC CAC AAA TAT GCT TTG GCA TTG GCC TTT TCA     261
Leu Lys Thr Thr Lys Asn His Lys Tyr Ala Leu Ala Leu Ala Phe Ser
35                  40                  45                  50

ATT GAT GAA ATC AAC AGG AAT CCT GAT CTT TTA CCA AAT ATG TCT TTG     309
Ile Asp Glu Ile Asn Arg Asn Pro Asp Leu Leu Pro Asn Met Ser Leu
                55                  60                  65
```

```
ATC ATA AAA TAC CCT TTG GGC CTT TGC GAT GGA CAA ACT ACA TTA CCT      357
Ile Ile Lys Tyr Pro Leu Gly Leu Cys Asp Gly Gln Thr Thr Leu Pro
            70                  75                  80

ACA CCC TAT TTA TTT AAT GAA ATA TAT TTT AGG CCT ATC CCT AAT TAT      405
Thr Pro Tyr Leu Phe Asn Glu Ile Tyr Phe Arg Pro Ile Pro Asn Tyr
        85                  90                  95

TTC TGT AAT GAA GAG ACT ATG TGT ACA TTT CTA CTT ACA GGA CCG CAT      453
Phe Cys Asn Glu Glu Thr Met Cys Thr Phe Leu Leu Thr Gly Pro His
    100                 105                 110

TGG ATA ACA TCT TAT AGT TTC TGG ATA CAC TTG AAC ATC TTC TTA TCT      501
Trp Ile Thr Ser Tyr Ser Phe Trp Ile His Leu Asn Ile Phe Leu Ser
115                 120                 125                 130

CCT AGT ATG AAC CCA AAG GAC ACA TCC CTA GCT TTG GCA ATG GTC TCC      549
Pro Ser Met Asn Pro Lys Asp Thr Ser Leu Ala Leu Ala Met Val Ser
                135                 140                 145

TTC TTA CTT TAT TTC AAG TGG AAC TGG GTC GGC CTT GTC ATC TCA GAT      597
Phe Leu Leu Tyr Phe Lys Trp Asn Trp Val Gly Leu Val Ile Ser Asp
            150                 155                 160

GAT GAT CAA GGC AAT CAA TTT CTC TCT GAG TTG AAA AAA GAG AGC AAA      645
Asp Asp Gln Gly Asn Gln Phe Leu Ser Glu Leu Lys Lys Glu Ser Lys
        165                 170                 175

ATC AAG GAA ATT TGC TTT GCA TTT GTG AGC ATG CTG GCA ATC GAT GAG      693
Ile Lys Glu Ile Cys Phe Ala Phe Val Ser Met Leu Ala Ile Asp Glu
    180                 185                 190

ATT TCA TTT TAT CAT AAA ACT GAA ATG TAC TAC AAC CAA ATT GTG ATG      741
Ile Ser Phe Tyr His Lys Thr Glu Met Tyr Tyr Asn Gln Ile Val Met
195                 200                 205                 210

TCA TCC ACA AAC GTT ATT ATC ATT TAT GGG AAA ACA GAG AGT ATT ATT      789
Ser Ser Thr Asn Val Ile Ile Ile Tyr Gly Lys Thr Glu Ser Ile Ile
                215                 220                 225

GAG TTG AGC TTC AGA ATG TGG GAA TCT CCA GTT ATC CAG AGA ATA TGG      837
Glu Leu Ser Phe Arg Met Trp Glu Ser Pro Val Ile Gln Arg Ile Trp
            230                 235                 240

GTC ACC ACA AAA GAA ATG AAT TTC CCT ACC AGT AAG AGA GAT TTA ACT      885
Val Thr Thr Lys Glu Met Asn Phe Pro Thr Ser Lys Arg Asp Leu Thr
        245                 250                 255

CAT GAC ACA TTC TAT GGG ACT CTT ACT TTT CTA CAC AGC CAT GGG GAG      933
His Asp Thr Phe Tyr Gly Thr Leu Thr Phe Leu His Ser His Gly Glu
    260                 265                 270

ATT TCA GGC TTT AAA AAT TTT GTA CAG ACA TGG TAC CAT CTT AGA ATC      981
Ile Ser Gly Phe Lys Asn Phe Val Gln Thr Trp Tyr His Leu Arg Ile
275                 280                 285                 290

ACT GAT TTG CAT CTA GTA ATG CCA GAG TGG AAA TAT TTT AAC TAT GAA     1029
Thr Asp Leu His Leu Val Met Pro Glu Trp Lys Tyr Phe Asn Tyr Glu
                295                 300                 305

GCC TCA GCA TCT AAC TGT AAA ATA TTG AAG AAC TAT TCA TCC AGT GCC     1077
Ala Ser Ala Ser Asn Cys Lys Ile Leu Lys Asn Tyr Ser Ser Ser Ala
            310                 315                 320

TCA TTG GAA TGG TTA ATG GAG CAG ACA TTT GAC ATG GTC TTT AGT GAT     1125
```

```
Ser Leu Glu Trp Leu Met Glu Gln Thr Phe Asp Met Val Phe Ser Asp
        325                 330                 335

GGA AGT CGG GAT ATA TAT AAT GCT GTA AAT GCC ATG GCC CAT GCA CTC    1173
Gly Ser Arg Asp Ile Tyr Asn Ala Val Asn Ala Met Ala His Ala Leu
    340                 345                 350

CAT GAG ATG AAT CTG CAC CTG GTT GAT AAT CAG GCA ATA GAC AAT GGG    1221
His Glu Met Asn Leu His Leu Val Asp Asn Gln Ala Ile Asp Asn Gly
355                 360                 365                 370

AAA GGA GCC AGT TCT CAC TGC TTT AAG ATA AAC TCC TTT CTC AGA AAG    1269
Lys Gly Ala Ser Ser His Cys Phe Lys Ile Asn Ser Phe Leu Arg Lys
                375                 380                 385

ACC CAC TTC ACT AAT CCT CTT GGG GAC AGA GTG ATT ATG AAA GAG AGA    1317
Thr His Phe Thr Asn Pro Leu Gly Asp Arg Val Ile Met Lys Glu Arg
            390                 395                 400

GAA ATA CTG CAA GAA GAC TAT AAC ATT TTT CAC ACT TGG AAT TTT TCT    1365
Glu Ile Leu Gln Glu Asp Tyr Asn Ile Phe His Thr Trp Asn Phe Ser
        405                 410                 415

CAG CAC ATT GGT TTT AAG GTG AAG ATA GGA AAG TTC AGC CCA TAT TTT    1413
Gln His Ile Gly Phe Lys Val Lys Ile Gly Lys Phe Ser Pro Tyr Phe
    420                 425                 430

CCA CAT GGC AGG CAC TTT CAC CTA TAT GTA GAC ATG ATT GAG TTG GCT    1461
Pro His Gly Arg His Phe His Leu Tyr Val Asp Met Ile Glu Leu Ala
435                 440                 445                 450

ACA GGA AGT AGA AAG ATG CCA TCC TCT GTG TGC ACT GAA GAT TGT AGT    1509
Thr Gly Ser Arg Lys Met Pro Ser Ser Val Cys Thr Glu Asp Cys Ser
                455                 460                 465

CCT GGA TAC AGA AGA TTC TGG AAG GAG GGA ATG GCA GCC TGC TGT TTT    1557
Pro Gly Tyr Arg Arg Phe Trp Lys Glu Gly Met Ala Ala Cys Cys Phe
            470                 475                 480

GTT TGC AGT CCC TGC CCT GAA AAT GCA ATT TCT AAT GAG ACA AAT ATG    1605
Val Cys Ser Pro Cys Pro Glu Asn Ala Ile Ser Asn Glu Thr Asn Met
        485                 490                 495

GAT CAG TGT GTG AAT TGT CCA GAA TAC CAA TAT GCC AAT ACA AAG CGG    1653
Asp Gln Cys Val Asn Cys Pro Glu Tyr Gln Tyr Ala Asn Thr Lys Arg
    500                 505                 510

GAC AAA TGC ATT CAG AAA AAT GTG ATG TTT CTA AGC TAC AAA GAC CCC    1701
Asp Lys Cys Ile Gln Lys Asn Val Met Phe Leu Ser Tyr Lys Asp Pro
515                 520                 525                 530

CTT GGG GAT GAC TCT TGC CTT CAT AGC CTT CTT TTT CTC TGC ATT AAC    1749
Leu Gly Asp Asp Ser Cys Leu His Ser Leu Leu Phe Leu Cys Ile Asn
                535                 540                 545

AGC TGT TGT ACT TAGGGTCTTT GTGAAGCACC ATGACACTCC TATTGTGAAG GCCAA  1806
Ser Cys Cys Thr
            550

TAACAGAATC CTCAGCTACC TATTAATCAC GTCTCTCTTG TTCTGTTTTC TCTGCTCATT   1866
TTTCTTCATT GGCCATCCTA ACAGAGCAAC CTGCATCTTA CAGCAAATCA CATTTGGAAT   1926
TGTATTCACT GTGGCTATTT CTACAATTTT GGCAAAAACA ATCACTGTGG TTCTGGCTTT   1986
CAAAGTCACA AACCCAGGAA GAAGGTTGAG AAACTTCCTA GTATTGGGTA CACTCAACTA   2046
CATTATCCCC ATATGTTCCC TGTTTCAATG TATTCTGTGT GCAATCTGGC TAGCAGTTTC   2106
TCCTCCCTTT GTTGATACTG ATGAACACAC TGAGTATGGC CACATCATCA TTGTGTGCAA   2166
```

```
CAAAGGCTCA GTAACTGCAT TCTACTGTGT CCTGGGATAC TTGGCCTGCT TGGCACTTGC   2226
AAGCTTCACT GTGGCTTTCT TGGCAAAGAA TCTGCCAGAC ACATTCAATG AAGCCAAGTT   2286
CTTGACCTTC AGCATGCTGG TGTTCTGCAG TGTCTGGGTC ACCTTCCTCC CTGTCTACCA   2346
CAGCACCAAG GGCAAGATCA TGGTTGCTGT GGAGATATTC TCCATTTTGG CATCCAGTGC   2406
AGGGATGCTT GGATGCATCT TTGCACCCAA GATTTACATC ATTTTAATGA GACCAGAGAG   2466
AAATGCTATC CAAAAGATCA GGGAGAAATC ATATTTCTGA ACAAATTATT TCAGAATTTC   2526
TATCAAATGT AAACATGGTA TATACCCATC AAATATTGTG TTACAGTGCA TGTATCTAGT   2586
TTTAGAATCA CTCTCACTGG TACCCCTAGT GATGTCTAGA AATATCATAT CTACCAATCT   2646
TGAATACATT GTCCATAAAA TCTTGTACAT ATTCACTAGC TTAGTTTCCT GTGGGAGAAC   2706
TAAAATTCTC AAATTATTAT TACAATTTTA TTCATAATTT TGCTCTCATG GCAAATCAGA   2766
ACTCATTTTC TAATTTCCAG TAACAACACA TACATGACAG AATACTGATT TTCAGCTATT   2826
CTTTAAGCTA TTGGCCAATA GACTAAGGTG GAAATGTTCT TTTTCTTTCT GAAACACAAA   2886
AATATTATAT CATATAATAC ACAGAAGTCA GGGACCCCTA TGGATGAATT AGGGAATAGT   2946
TGGAAGAAGC TGGCTGAGTA GAAGGGTGAC CCATAGGAAG ACCAGCAGTC TCACCTAACA   3006
AGGACAACCA AGATCTTGCT GACACTGAAT CACTTGCTAG GCAGTTGATT TGAGGCCCCT   3066
GACACATATC AAGCATAGGA CTACATTGGC TGGCCTCAGT GGGAGAAGAC AACCTAACCC   3126
CCTAGAGACT TGAGGCCCCA GGCTAAGGGG AGGTTGGGGG TTTTGAAAGT TGGGGATATT   3186
ATCTTGGAGT TGGGGAGGGG TATGGGATGA AGAAGAGTCA GGAGGCAGGT GCTGGTTGGA   3246
GTATAATGAC TGGACTGTAA ATAAAAGACT AACAACCAAA AATAAATAAA ATAACTTAAA   3306
A                                                                   3307
```

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 550 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

```
Met Lys Leu Arg Asp Lys Asp Leu Ser Ile Thr Cys Ser Phe Ile Leu
 1               5                  10                  15
Glu Ala Val Gln Met Pro Thr Glu Asn Asp Tyr Phe Asn Gln Thr Leu
            20                  25                  30
Asn Ile Leu Lys Thr Thr Lys Asn His Lys Tyr Ala Leu Ala Leu Ala
        35                  40                  45
Phe Ser Ile Asp Glu Ile Asn Arg Asn Pro Asp Leu Leu Pro Asn Met
    50                  55                  60
Ser Leu Ile Ile Lys Tyr Pro Leu Gly Leu Cys Asp Gly Gln Thr Thr
65                  70                  75                  80
Leu Pro Thr Pro Tyr Leu Phe Asn Glu Ile Tyr Phe Arg Pro Ile Pro
                85                  90                  95
Asn Tyr Phe Cys Asn Glu Glu Thr Met Cys Thr Phe Leu Leu Thr Gly
            100                 105                 110
Pro His Trp Ile Thr Ser Tyr Ser Phe Trp Ile His Leu Asn Ile Phe
        115                 120                 125
Leu Ser Pro Ser Met Asn Pro Lys Asp Thr Ser Leu Ala Leu Ala Met
    130                 135                 140
Val Ser Phe Leu Leu Tyr Phe Lys Trp Asn Trp Val Gly Leu Val Ile
145                 150                 155                 160
Ser Asp Asp Asp Gln Gly Asn Gln Phe Leu Ser Glu Leu Lys Lys Glu
                165                 170                 175
Ser Lys Ile Lys Glu Ile Cys Phe Ala Phe Val Ser Met Leu Ala Ile
            180                 185                 190
Asp Glu Ile Ser Phe Tyr His Lys Thr Glu Met Tyr Tyr Asn Gln Ile
        195                 200                 205
Val Met Ser Ser Thr Asn Val Ile Ile Ile Tyr Gly Lys Thr Glu Ser
    210                 215                 220
Ile Ile Glu Leu Ser Phe Arg Met Trp Glu Ser Pro Val Ile Gln Arg
225                 230                 235                 240
Ile Trp Val Thr Thr Lys Glu Met Asn Phe Pro Thr Ser Lys Arg Asp
```

```
                245                 250                 255
Leu Thr His Asp Thr Phe Tyr Gly Thr Leu Thr Phe Leu His Ser His
            260                 265                 270
Gly Glu Ile Ser Gly Phe Lys Asn Phe Val Gln Thr Trp Tyr His Leu
        275                 280                 285
Arg Ile Thr Asp Leu His Leu Val Met Pro Glu Trp Lys Tyr Phe Asn
    290                 295                 300
Tyr Glu Ala Ser Ala Ser Asn Cys Lys Ile Leu Lys Asn Tyr Ser Ser
305                 310                 315                 320
Ser Ala Ser Leu Glu Trp Leu Met Glu Gln Thr Phe Asp Met Val Phe
                325                 330                 335
Ser Asp Gly Ser Arg Asp Ile Tyr Asn Ala Val Asn Ala Met Ala His
            340                 345                 350
Ala Leu His Glu Met Asn Leu His Leu Val Asp Asn Gln Ala Ile Asp
        355                 360                 365
Asn Gly Lys Gly Ala Ser Ser His Cys Phe Lys Ile Asn Ser Phe Leu
    370                 375                 380
Arg Lys Thr His Phe Thr Asn Pro Leu Gly Asp Arg Val Ile Met Lys
385                 390                 395                 400
Glu Arg Glu Ile Leu Gln Glu Asp Tyr Asn Ile Phe His Thr Trp Asn
                405                 410                 415
Phe Ser Gln His Ile Gly Phe Lys Val Lys Ile Gly Lys Phe Ser Pro
            420                 425                 430
Tyr Phe Pro His Gly Arg His Phe His Leu Tyr Val Asp Met Ile Glu
        435                 440                 445
Leu Ala Thr Gly Ser Arg Lys Met Pro Ser Ser Val Cys Thr Glu Asp
    450                 455                 460
Cys Ser Pro Gly Tyr Arg Arg Phe Trp Lys Glu Gly Met Ala Ala Cys
465                 470                 475                 480
Cys Phe Val Cys Ser Pro Cys Pro Glu Asn Ala Ile Ser Asn Glu Thr
                485                 490                 495
Asn Met Asp Gln Cys Val Asn Cys Pro Glu Tyr Gln Tyr Ala Asn Thr
            500                 505                 510
Lys Arg Asp Lys Cys Ile Gln Lys Asn Val Met Phe Leu Ser Tyr Lys
        515                 520                 525
Asp Pro Leu Gly Asp Asp Ser Cys Leu His Ser Leu Leu Phe Leu Cys
    530                 535                 540
Ile Asn Ser Cys Cys Thr
545                 550
```

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3938 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 46...2424
(D) OTHER INFORMATION: GoVN7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

```
CGGCACGAGC CCAGGTTTAA GGCTGGAAAA AATATGTTCA TTTTG ATG ATA GTA TTC    57
                                                  Met Ile Val Phe
                                                   1

TTT CTC CTC AAC ATT CCA CTT CTC ATG GCA AAT TCC GTT GAT CCC AGG     105
Phe Leu Leu Asn Ile Pro Leu Leu Met Ala Asn Ser Val Asp Pro Arg
 5                  10                  15                  20
```

```
TGC TTT TGG AAA ATA AAT TTG AAT GAA GTC AAG GAT ATA GAT TTA GAT      153
Cys Phe Trp Lys Ile Asn Leu Asn Glu Val Lys Asp Ile Asp Leu Asp
                25                  30                  35

ACA AGT TGT TAC TTC ATC CTT GAG GCA GTT CAG TTG CCT ATG GAG AAA      201
Thr Ser Cys Tyr Phe Ile Leu Glu Ala Val Gln Leu Pro Met Glu Lys
            40                  45                  50

GAT TAT TTC AAC CAG ACT CTG AAT GTC CTA AAA ACA ACC AAA TAC AAC      249
Asp Tyr Phe Asn Gln Thr Leu Asn Val Leu Lys Thr Thr Lys Tyr Asn
        55                  60                  65

AGA TAT GCA TTG GCA TTA GCC TTT ACA ATG GAT GAA ATA AAC AGG AAT      297
Arg Tyr Ala Leu Ala Leu Ala Phe Thr Met Asp Glu Ile Asn Arg Asn
    70                  75                  80

CCT CAT ATT TTA CCA AAC ATG TCT TTG ATT ATA AAA CAT ACA TTG GGC      345
Pro His Ile Leu Pro Asn Met Ser Leu Ile Ile Lys His Thr Leu Gly
85                  90                  95                  100

CAC TGT GAT GGA AAT ATC CCA CTC CGC TTA CTT AAT CAA ATA TTT TAT      393
His Cys Asp Gly Asn Ile Pro Leu Arg Leu Leu Asn Gln Ile Phe Tyr
                105                 110                 115

ATG CCT TTT CCT AAT TAT GGC TGT AAT GAA GAG ACT ATG TGT TCA TTT      441
Met Pro Phe Pro Asn Tyr Gly Cys Asn Glu Glu Thr Met Cys Ser Phe
            120                 125                 130

ATG CTT ATG GGA CCG AAT TTG TGG CCA TCT GTA GAT TTT TTC ATT CAC      489
Met Leu Met Gly Pro Asn Leu Trp Pro Ser Val Asp Phe Phe Ile His
        135                 140                 145

TTG AAC ATC TTA TTT CCT CAT TTC CTT CAG ATT TCC TTC GGA CCT TTC      537
Leu Asn Ile Leu Phe Pro His Phe Leu Gln Ile Ser Phe Gly Pro Phe
    150                 155                 160

CAT TCC ATT TTC AGT GAT AAT GAA CAA TTT CCT TAT ATC TAT CAG ATG      585
His Ser Ile Phe Ser Asp Asn Glu Gln Phe Pro Tyr Ile Tyr Gln Met
165                 170                 175                 180

ACC CCA AAG GAT ACA TCA CTA GCA TTG GCA ATG GTC TCT TTC ATA CTT      633
Thr Pro Lys Asp Thr Ser Leu Ala Leu Ala Met Val Ser Phe Ile Leu
                185                 190                 195

TAC TTC AAC TGG AAC TGG GTT GGT CTT GTC CTC TCA GAT AAT GAT GAA      681
Tyr Phe Asn Trp Asn Trp Val Gly Leu Val Leu Ser Asp Asn Asp Glu
            200                 205                 210

GGC AAT CAA TTT CTC ACA GAG TTG AAA AAA GAG ACC CAC AAC ACG GAA      729
Gly Asn Gln Phe Leu Thr Glu Leu Lys Lys Glu Thr His Asn Thr Glu
        215                 220                 225

ATA TGC TTT GCC TTT GTG AAC ATG ATG GCA ATC AAT GAG AAT TCA TCC      777
Ile Cys Phe Ala Phe Val Asn Met Met Ala Ile Asn Glu Asn Ser Ser
    230                 235                 240

ATG AAA AAA ACT GAC ATG TAC TAC AAC CAA ATT GTG ATG TCA ACC GCA      825
Met Lys Lys Thr Asp Met Tyr Tyr Asn Gln Ile Val Met Ser Thr Ala
245                 250                 255                 260

AAT GTT ATT ATC ATT TAT GGG GAA CGA CCC AGT ATT ATT GAA CTG TGT      873
Asn Val Ile Ile Ile Tyr Gly Glu Arg Pro Ser Ile Ile Glu Leu Cys
                265                 270                 275

TTC AGA ACA TGG ACA TCT CCA GTC ATA CAG AGG ATA TGG GTT ACC AAA      921
```

```
Phe Arg Thr Trp Thr Ser Pro Val Ile Gln Arg Ile Trp Val Thr Lys
            280                 285                 290

TCA GAG TTG TAT TTC CCA ACA AGT AAG AGA GAC TTA AGT CAT GGA ACA      969
Ser Glu Leu Tyr Phe Pro Thr Ser Lys Arg Asp Leu Ser His Gly Thr
        295                 300                 305

TTC TAT GGA ACT CTA GCA TTT CAA CAA CAC CAT GAT GTG ATT TCT GGA     1017
Phe Tyr Gly Thr Leu Ala Phe Gln Gln His His Asp Val Ile Ser Gly
    310                 315                 320

TTT AAA AAT TTT GTA CAG ACA TGG TAC CAT CTC AAA AGC ATG GAT TTA     1065
Phe Lys Asn Phe Val Gln Thr Trp Tyr His Leu Lys Ser Met Asp Leu
325                 330                 335                 340

TAT TTA TTA AAG CCA GAG TGG GGT TTC TTT GAA TAT GAA ACC TCA GCA     1113
Tyr Leu Leu Lys Pro Glu Trp Gly Phe Phe Glu Tyr Glu Thr Ser Ala
                345                 350                 355

TCT TAC TGT AAA ATA CTG ATG AGT AAT TCA TCG AAT GTC TCA TTG GAA     1161
Ser Tyr Cys Lys Ile Leu Met Ser Asn Ser Ser Asn Val Ser Leu Glu
            360                 365                 370

TGG CTA ATG GAA CAG AAG TTT GAC ATA GCC TTT AAT GAC AAT AGT CAT     1209
Trp Leu Met Glu Gln Lys Phe Asp Ile Ala Phe Asn Asp Asn Ser His
        375                 380                 385

AGT ATA TAC AAT GCT GTG TAC GCC ATG GCC CAT GCT CTC CAT GAA AAG     1257
Ser Ile Tyr Asn Ala Val Tyr Ala Met Ala His Ala Leu His Glu Lys
    390                 395                 400

AAT CTG AAA CAA ATT GAT AAT CAG GAA ATC AGC TAT GGC AAA GGA GCA     1305
Asn Leu Lys Gln Ile Asp Asn Gln Glu Ile Ser Tyr Gly Lys Gly Ala
405                 410                 415                 420

AGT ACT CAC TGC TTG AAG TTA CAC TCA TTT TTG AGA ACG ATC CAC TTC     1353
Ser Thr His Cys Leu Lys Leu His Ser Phe Leu Arg Thr Ile His Phe
                425                 430                 435

ACC AAT CCT TTT GGG GAG AGA GTG ATT ATG AAA GAG AGA GTA AGA GTG     1401
Thr Asn Pro Phe Gly Glu Arg Val Ile Met Lys Glu Arg Val Arg Val
            440                 445                 450

CAG GAA GAC TAT GAC ATT GTT CAC CTG CAG AAC TGC TCA CAA CAC CTT     1449
Gln Glu Asp Tyr Asp Ile Val His Leu Gln Asn Cys Ser Gln His Leu
        455                 460                 465

AGG ATT AAG GTG AAG ATA GGG CAG TTC AGC CCA TAT TTT CCA CAT GGT     1497
Arg Ile Lys Val Lys Ile Gly Gln Phe Ser Pro Tyr Phe Pro His Gly
    470                 475                 480

GGA CAA TTT CAC TTA TAT GAA GAC ATG ATT GAT TTG GCC ACA GGA AGT     1545
Gly Gln Phe His Leu Tyr Glu Asp Met Ile Asp Leu Ala Thr Gly Ser
485                 490                 495                 500

AGA AAG ATG CCT TTA TCT ATG TGT AGT GCA GAT TGT CGT CCT GGA TAC     1593
Arg Lys Met Pro Leu Ser Met Cys Ser Ala Asp Cys Arg Pro Gly Tyr
                505                 510                 515

AGA AAA TTC TGG AAG GAG GGA ATG GCA GCC TGC TGT TTT GTT TGC AGT     1641
Arg Lys Phe Trp Lys Glu Gly Met Ala Ala Cys Cys Phe Val Cys Ser
            520                 525                 530

CCC TGT CCA GAC AAT GAA ATT TCT AAT GAA ACA ACT GTG GTA CTT TGG     1689
Pro Cys Pro Asp Asn Glu Ile Ser Asn Glu Thr Thr Val Val Leu Trp
```

```
        535                 540                 545

GTC TTT GTG AAG CAC CAT GAC ACT CCT ATT GTG AAG GCC AAT AAC AGA     1737
Val Phe Val Lys His His Asp Thr Pro Ile Val Lys Ala Asn Asn Arg
    550                 555                 560

ATC CTC AGC TAC ATA TTA ATC ATG TCA CTC ATG TTC TGC TTT CTG TGC     1785
Ile Leu Ser Tyr Ile Leu Ile Met Ser Leu Met Phe Cys Phe Leu Cys
565                 570                 575                 580

TCC TTT TTC TTC ATT GGC CAT CCT AAC AGA GGT ACC TGT ATC TTA CAG     1833
Ser Phe Phe Phe Ile Gly His Pro Asn Arg Gly Thr Cys Ile Leu Gln
                585                 590                 595

CAA ATC ACA TTT GGA ATT GTA TTC ACT GTG GCT GTT TCC ACA GTT CTG     1881
Gln Ile Thr Phe Gly Ile Val Phe Thr Val Ala Val Ser Thr Val Leu
            600                 605                 610

GCC AAA ACA ATC ACT GTG CTT CTG GCT TTT CAA GTC ACA GAC ACA GGA     1929
Ala Lys Thr Ile Thr Val Leu Leu Ala Phe Gln Val Thr Asp Thr Gly
        615                 620                 625

AGA AAG TTA AGA AAC TTC CTG GTA TCG GGG ACA CCC AAC TAC ATT ATT     1977
Arg Lys Leu Arg Asn Phe Leu Val Ser Gly Thr Pro Asn Tyr Ile Ile
    630                 635                 640

CCC ATA TGT TCC CTG TTG CAA TGC ACT CTG TGT GCA ATT TGG CTA GCA     2025
Pro Ile Cys Ser Leu Leu Gln Cys Thr Leu Cys Ala Ile Trp Leu Ala
645                 650                 655                 660

GTT TCT CCA CCA TTT GTT GAT ATC GAT GAA CAT TCT GAG CAT GGT CAC     2073
Val Ser Pro Pro Phe Val Asp Ile Asp Glu His Ser Glu His Gly His
                665                 670                 675

ATC ATA ATT GTG TGC AAC AAG GGA TCT GTT ATG GCA TTC TAC TGT GTC     2121
Ile Ile Ile Val Cys Asn Lys Gly Ser Val Met Ala Phe Tyr Cys Val
            680                 685                 690

CTG GGA TAT TTG GCC TTC CTG GCC CTT GGA AGT TTC ACG ATG GCT TTC     2169
Leu Gly Tyr Leu Ala Phe Leu Ala Leu Gly Ser Phe Thr Met Ala Phe
        695                 700                 705

TTG GCA AAG AAT CTG CCT GAC ACA TTC AAT GAA GCC AAG TTC TTG ACC     2217
Leu Ala Lys Asn Leu Pro Asp Thr Phe Asn Glu Ala Lys Phe Leu Thr
    710                 715                 720

TTC AGC ATG CTA GTG TTC TGC AGT GTC TGG ATC ACG TTC CTT CCT GTC     2265
Phe Ser Met Leu Val Phe Cys Ser Val Trp Ile Thr Phe Leu Pro Val
725                 730                 735                 740

TAC CAT AGC ACC AAG GGC AGA GTC ATG GTT GCT GTT GAA ATT TTC TCC     2313
Tyr His Ser Thr Lys Gly Arg Val Met Val Ala Val Glu Ile Phe Ser
                745                 750                 755

ATT TTG ACA TCC AGT GCA GGG ATG CTT GGA TGC GTC TTT GCA CCC AAA     2361
Ile Leu Thr Ser Ser Ala Gly Met Leu Gly Cys Val Phe Ala Pro Lys
            760                 765                 770

ATT TAC ATC ATT TTA ATG AAA CCA GAG AGA ATT CTA TCC AAA AGA CAG     2409
Ile Tyr Ile Ile Leu Met Lys Pro Glu Arg Ile Leu Ser Lys Arg Gln
        775                 780                 785

GAG AAA TCA CGT TTC TAAACAGATA TTTTAGAAAT TCTGTCAAAT GTACAGTTGT T   2465
Glu Lys Ser Arg Phe
    790
```

```
ATATACCCAC CAAATATTTG GTTACAGTGC ATAAATCTAG TTTTAGAACT CTCACTAGTT   2525
CCTCTAATGA TATCTAGAAA TATTGTATCT ACCAATCTTA CATTCATTAT CCATAAAATC   2585
CTGCACTCAT TCACTTGTTT GTTCTACTCT GTGAGAAATA TAATTCCCAA TGTAGTATTA   2645
AATTTTTTCT AAAAATTTTG CTTTAATTGA CATTTTTTCC CTTATAACTT CAAGTACATT   2705
TGATAAGGCA TTTGAATCTA TAACCTTTTA TACAATAAGA TCCAGGACAG ACAGGATTAC   2765
ACATAGAAAC CGTCTATCGA ATCAAACAAT CAATCAGACT AAAAAACAAA GAATCAACAA   2825
AGATAACATC AGAATACATT ATCTGATTTC CAGTAGAAGC ACATATGTGA CAGAATACTG   2885
TCTGTTTTTA TAGTTCCTCT TCAAGCTATT GTATTGGTCA GCAGTCTAAG GTAGAAGTTT   2945
TTTTGTCACA AACACAAAAA TATTGTATCC AACAATGGAC AGAATCCAGT GAGCACCCTG   3005
TTCAAATTTG GAGATAGTTG GAATATCATG AAAAAGAGGG TGACCCATAA GAATACCAGC   3065
ATTCTCAACT AACCTGGACA ACCACGAATT TGAGCTGCTG ACCAGGCAGC ATACATAAGC   3125
TGATATGAGG CTCCCAGCAC AGATGCAACA TAGGGCTGCC TGGTCTGGCC TCAGTGGAAG   3185
AAGACACATT TAAACCACAA GAGACAGGAG TCACAAGGGA TTGGGAAGGT GTGATGGTTT   3245
GCATATGCTT GGCTCAGGAA GTGGCACTAT TAGAAGGTGT AGACTTGATG GAGGAATTTG   3305
TCACTGTAGG GGTGGGCTTG GAGATCCACC TCATAGCTGC CTGGGGATGC TCAGTCTGTT   3365
CCTGGCTTCC TTCAGGTGAA GATATAGAAC TCAGATCCTC CTTCACCAAG CCTGCCTGGA   3425
TGCTGTGATG CTGCCATGCT CCGACCTTGA TGATAATGGA CTGAACCTCT GAACATGTAA   3485
GCTGGCTCCA ATTAAAGGTT GTCCTTTATA AAACTTCCAT TGATCACAGT GTCTGTACAT   3545
AGCAATAAGA CCCAAACTAA GACAGAAGGT GTGTGGATTG GGGAAGTGGG GATTTCCTCT   3605
TGGAGGTGGG GAAGTAGTCA AAGATTAAAT TGGGAAGGGG ATAATGAGTA CACCGTAAAA   3665
AGTATTAAAG AATAAAATAC TAAAAAATTA ATTAAATAGG ATTGTGAATA TATTAACATG   3725
CTATTATATT ATAGTTCTGG AAGGGATAGG TAAAACTCCT GATGGTGGTT TGTACCTAAT   3785
TTTTCTTAGA GCTTGCCCTT TGTATTCAGT TGTGATTGAA ATCCTGGGCT CACAAAATTC   3845
TAGTACTATG GATATGGAGG CAGATACTTT GATTACGCTG CTTCCTAGAA ATAAATTTTC   3905
CAAAAACCAA AAAAAAAAAA AAAAAAAAAA AAA                                3938
```

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 793 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

```
Met Ile Val Phe Phe Leu Leu Asn Ile Pro Leu Leu Met Ala Asn Ser
 1               5                  10                  15
Val Asp Pro Arg Cys Phe Trp Lys Ile Asn Leu Asn Glu Val Lys Asp
            20                  25                  30
Ile Asp Leu Asp Thr Ser Cys Tyr Phe Ile Leu Glu Ala Val Gln Leu
        35                  40                  45
Pro Met Glu Lys Asp Tyr Phe Asn Gln Thr Leu Asn Val Leu Lys Thr
    50                  55                  60
Thr Lys Tyr Asn Arg Tyr Ala Leu Ala Leu Ala Phe Thr Met Asp Glu
65                  70                  75                  80
Ile Asn Arg Asn Pro His Ile Leu Pro Asn Met Ser Leu Ile Ile Lys
                85                  90                  95
His Thr Leu Gly His Cys Asp Gly Asn Ile Pro Leu Arg Leu Leu Asn
            100                 105                 110
Gln Ile Phe Tyr Met Pro Phe Pro Asn Tyr Gly Cys Asn Glu Glu Thr
        115                 120                 125
Met Cys Ser Phe Met Leu Met Gly Pro Asn Leu Trp Pro Ser Val Asp
    130                 135                 140
Phe Phe Ile His Leu Asn Ile Leu Phe Pro His Phe Leu Gln Ile Ser
145                 150                 155                 160
Phe Gly Pro Phe His Ser Ile Phe Ser Asp Asn Glu Gln Phe Pro Tyr
                165                 170                 175
Ile Tyr Gln Met Thr Pro Lys Asp Thr Ser Leu Ala Leu Ala Met Val
            180                 185                 190
Ser Phe Ile Leu Tyr Phe Asn Trp Asn Trp Val Gly Leu Val Leu Ser
        195                 200                 205
```

```
Asp Asn Asp Glu Gly Asn Gln Phe Leu Thr Glu Leu Lys Lys Glu Thr
    210                 215                 220
His Asn Thr Glu Ile Cys Phe Ala Phe Val Asn Met Met Ala Ile Asn
225                 230                 235                 240
Glu Asn Ser Ser Met Lys Lys Thr Asp Met Tyr Tyr Asn Gln Ile Val
                245                 250                 255
Met Ser Thr Ala Asn Val Ile Ile Ile Tyr Gly Glu Arg Pro Ser Ile
            260                 265                 270
Ile Glu Leu Cys Phe Arg Thr Trp Thr Ser Pro Val Ile Gln Arg Ile
        275                 280                 285
Trp Val Thr Lys Ser Glu Leu Tyr Phe Pro Thr Ser Lys Arg Asp Leu
    290                 295                 300
Ser His Gly Thr Phe Tyr Gly Thr Leu Ala Phe Gln Gln His His Asp
305                 310                 315                 320
Val Ile Ser Gly Phe Lys Asn Phe Val Gln Thr Trp Tyr His Leu Lys
                325                 330                 335
Ser Met Asp Leu Tyr Leu Leu Lys Pro Glu Trp Gly Phe Phe Glu Tyr
            340                 345                 350
Glu Thr Ser Ala Ser Tyr Cys Lys Ile Leu Met Ser Asn Ser Ser Asn
        355                 360                 365
Val Ser Leu Glu Trp Leu Met Glu Gln Lys Phe Asp Ile Ala Phe Asn
    370                 375                 380
Asp Asn Ser His Ser Ile Tyr Asn Ala Val Tyr Ala Met Ala His Ala
385                 390                 395                 400
Leu His Glu Lys Asn Leu Lys Gln Ile Asp Asn Gln Glu Ile Ser Tyr
                405                 410                 415
Gly Lys Gly Ala Ser Thr His Cys Leu Lys Leu His Ser Phe Leu Arg
            420                 425                 430
Thr Ile His Phe Thr Asn Pro Phe Gly Glu Arg Val Ile Met Lys Glu
        435                 440                 445
Arg Val Arg Val Gln Glu Asp Tyr Asp Ile Val His Leu Gln Asn Cys
    450                 455                 460
Ser Gln His Leu Arg Ile Lys Val Lys Ile Gly Gln Phe Ser Pro Tyr
465                 470                 475                 480
Phe Pro His Gly Gly Gln Phe His Leu Tyr Glu Asp Met Ile Asp Leu
                485                 490                 495
Ala Thr Gly Ser Arg Lys Met Pro Leu Ser Met Cys Ser Ala Asp Cys
            500                 505                 510
Arg Pro Gly Tyr Arg Lys Phe Trp Lys Glu Gly Met Ala Ala Cys Cys
        515                 520                 525
Phe Val Cys Ser Pro Cys Pro Asp Asn Glu Ile Ser Asn Glu Thr Thr
    530                 535                 540
Val Val Leu Trp Val Phe Val Lys His His Asp Thr Pro Ile Val Lys
545                 550                 555                 560
Ala Asn Asn Arg Ile Leu Ser Tyr Ile Leu Ile Met Ser Leu Met Phe
                565                 570                 575
Cys Phe Leu Cys Ser Phe Phe Phe Ile Gly His Pro Asn Arg Gly Thr
            580                 585                 590
Cys Ile Leu Gln Gln Ile Thr Phe Gly Ile Val Phe Thr Val Ala Val
        595                 600                 605
Ser Thr Val Leu Ala Lys Thr Ile Thr Val Leu Leu Ala Phe Gln Val
    610                 615                 620
Thr Asp Thr Gly Arg Lys Leu Arg Asn Phe Leu Val Ser Gly Thr Pro
625                 630                 635                 640
Asn Tyr Ile Ile Pro Ile Cys Ser Leu Leu Gln Cys Thr Leu Cys Ala
                645                 650                 655
Ile Trp Leu Ala Val Ser Pro Pro Phe Val Asp Ile Asp Glu His Ser
            660                 665                 670
Glu His Gly His Ile Ile Ile Val Cys Asn Lys Gly Ser Val Met Ala
        675                 680                 685
Phe Tyr Cys Val Leu Gly Tyr Leu Ala Phe Leu Ala Leu Gly Ser Phe
    690                 695                 700
Thr Met Ala Phe Leu Ala Lys Asn Leu Pro Asp Thr Phe Asn Glu Ala
705                 710                 715                 720
Lys Phe Leu Thr Phe Ser Met Leu Val Phe Cys Ser Val Trp Ile Thr
```

```
                725                 730                 735
Phe Leu Pro Val Tyr His Ser Thr Lys Gly Arg Val Met Val Ala Val
            740                 745                 750
Glu Ile Phe Ser Ile Leu Thr Ser Ser Ala Gly Met Leu Gly Cys Val
        755                 760                 765
Phe Ala Pro Lys Ile Tyr Ile Ile Leu Met Lys Pro Glu Arg Ile Leu
    770                 775                 780
Ser Lys Arg Gln Glu Lys Ser Arg Phe
785                 790
```

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3359 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 59...2452
(D) OTHER INFORMATION: GoVN13C

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

```
CGGCACGAGC ACAGTCCACT CTGTCAGGGT TTAAGGCAGG AAAAACATGC TCATTTTG AT     60
                                                                 Met
                                                                  1

GGT AAT ATT CTT CCT TCT CAA CAT TCC ATT TCT CCT GGC AAA TTT CAT      108
 Val Ile Phe Phe Leu Leu Asn Ile Pro Phe Leu Leu Ala Asn Phe Met
               5                  10                  15

GGA TCC CAG ATG CTT TTG GAA AAT AAA TTT GAA TGA AAT CAA GGA TGA      156
 Asp Pro Arg Cys Phe Trp Lys Ile Asn Leu Asn Glu Ile Lys Asp Glu
         20                  25                  30

AGT CCT TGG GAT GAC TTG TTC CTT CAT CCT TGA AAC AGT TCA GAA GAC      204
 Val Leu Gly Met Thr Cys Ser Phe Ile Leu Glu Thr Val Gln Lys Thr
     35                  40                  45

TAT GGA CAA AGA TTA TTT CAA CCA GAC TCT GAA TGT CCT AAA TAC AAC      252
 Met Asp Lys Asp Tyr Phe Asn Gln Thr Leu Asn Val Leu Asn Thr Thr
 50                  55                  60                  65

TAC AAA CCA CAA ATA TGC CTT GGC ATT GGC CTT TAC AGT GGA TGA AAT      300
 Thr Asn His Lys Tyr Ala Leu Ala Leu Ala Phe Thr Val Asp Glu Ile
                 70                  75                  80

CAA CAG GAA TCC TGA TCT TTT ACC AAA TAT GTC TCT GAT TAT AAA ATA      348
 Asn Arg Asn Pro Asp Leu Leu Pro Asn Met Ser Leu Ile Ile Lys Tyr
             85                  90                  95

CAA TTT GGG TCA TTG TGA TGG AAA AAC TGT AAC AAC TCT ATC CGA TTT      396
 Asn Leu Gly His Cys Asp Gly Lys Thr Val Thr Thr Leu Ser Asp Leu
         100                 105                 110

ATT TAA TCC AAA TAA TCA TCT CCA TTT CCC CAA TTA TTT ATG TAA TGA      444
 Phe Asn Pro Asn Asn His Leu His Phe Pro Asn Tyr Leu Cys Asn Glu
     115                 120                 125

AGG GAT TAT GTG TTT GGT TCT GCT TAC AGG ACC ACA TTG GAG AGC ATC .    492
```

Gly Ile Met Cys Leu Val Leu Leu Thr Gly Pro His Trp Arg Ala Ser
130 135 140 14

TTT ATA TCT CTG GAT ATC CGT GTA TGT CTA CCT GTC TCC ACA TTT CCT 540
Leu Tyr Leu Trp Ile Ser Val Tyr Val Tyr Leu Ser Pro His Phe Leu
150 155 160

TCA GCT TTC CTA TGG ACC TTT CTA CTC CAT CTT CAG TGA TAA TGA ACA 588
Gln Leu Ser Tyr Gly Pro Phe Tyr Ser Ile Phe Ser Asp Asn Glu Gln
165 170 175

ATA TCC TTA TCT CTA TCA GAT GGG CCC AAA GGA CTC ATC ACT AGC ATT 636
Tyr Pro Tyr Leu Tyr Gln Met Gly Pro Lys Asp Ser Ser Leu Ala Leu
180 185 190

GGC AAT GGT CTC CTT CAT AAT TTA CTT CAA GTG GAA CTG GGT TGG GCT 684
Ala Met Val Ser Phe Ile Ile Tyr Phe Lys Trp Asn Trp Val Gly Leu
195 200 205

ATT TAT CTC AGA TGA TGA TCA AGG CAA TCA ATT TCT CTC AGA GTT GAA 732
Phe Ile Ser Asp Asp Asp Gln Gly Asn Gln Phe Leu Ser Glu Leu Lys
210 215 220 22

AAA AGA GAG CCA AAC CAA GGA TAT TTG CTT TGC CTT TGT GAA CAT GAT 780
Lys Glu Ser Gln Thr Lys Asp Ile Cys Phe Ala Phe Val Asn Met Ile
230 235 240

ATC AGT CAG TGA TGT TTC ATA CTA TCA TAA AAC TGA AAT GTA CTA CAA 828
Ser Val Ser Asp Val Ser Tyr Tyr His Lys Thr Glu Met Tyr Tyr Asn
245 250 255

CCA AAT TGT GAT GTC ATC CAC AAA GGT TAT TAT CAT TTA TGG GGA AAC 876
Gln Ile Val Met Ser Ser Thr Lys Val Ile Ile Ile Tyr Gly Glu Thr
260 265 270

AAA CAG TAT TAT TGA ATT GAG CTT CAG AAT GTG GTC ATC TCC AGT TAA 924
Asn Ser Ile Ile Glu Leu Ser Phe Arg Met Trp Ser Ser Pro Val Lys
275 280 285

ACA GAG AAT ATG GGT CAC CAC AAA ACA ATT TGA TTG CCC TAC CAG TAA 972
Gln Arg Ile Trp Val Thr Thr Lys Gln Phe Asp Cys Pro Thr Ser Lys
290 295 300 30

GAG AGA CTT AAC TCA TGG CAC ATT CTA TGG GAC CCT TAC ATT TCT ACA 1020
Arg Asp Leu Thr His Gly Thr Phe Tyr Gly Thr Leu Thr Phe Leu His
310 315 320

CCA CTA TGG TGA GAT TTC TGG CTT TAA AAA TTT TGT ACA GAC ACG GTA 1068
His Tyr Gly Glu Ile Ser Gly Phe Lys Asn Phe Val Gln Thr Arg Tyr
325 330 335

CAA TCT CAG AAG CAC AGA TTT ATA TCT AGT AAT GCC AGA GTG GAA ATA 1116
Asn Leu Arg Ser Thr Asp Leu Tyr Leu Val Met Pro Glu Trp Lys Tyr
340 345 350

TTT TAA CTA TGA AGC CTC AGC ATC TAA CTG TAA AAT ACT GAG AAA CTA 1164
Phe Asn Tyr Glu Ala Ser Ala Ser Asn Cys Lys Ile Leu Arg Asn Tyr
355 360 365

TTT ATC CAA TAT CTC ACT GGA ATG GCT AAT GGA ACA GAA ATT TGA CAT 1212
Leu Ser Asn Ile Ser Leu Glu Trp Leu Met Glu Gln Lys Phe Asp Met
370 375 380 38

GTC ATT TAG TGA TTA TAG TCA CAA CAT ATA CAA TGC TGT ATA TGC CAT 1260
Ser Phe Ser Asp Tyr Ser His Asn Ile Tyr Asn Ala Val Tyr Ala Ile

```
                390                 395                 400

TGC TCA TGC ACT CCA TGA GAA GAA TCT GCA AGA AGT TGA AAA TCA GGC   1308
 Ala His Ala Leu His Glu Lys Asn Leu Gln Glu Val Glu Asn Gln Ala
            405                 410                 415

AAT AAA CAA TGC GAA AGG AGA AAA TAC TCA CTG CTT GAA GCT AAA CTC   1356
 Ile Asn Asn Ala Lys Gly Glu Asn Thr His Cys Leu Lys Leu Asn Ser
        420                 425                 430

ATT TCT GAG AAA GAC CCA CTT CAC TAA TTC TCT TGG GAA CAG AGT AAT   1404
 Phe Leu Arg Lys Thr His Phe Thr Asn Ser Leu Gly Asn Arg Val Ile
    435                 440                 445

TAT GAA ACA GAG AGA AGT AGT GCA TGG AGA CTA TAA TAT TGT TCA CAT   1452
 Met Lys Gln Arg Glu Val Val His Gly Asp Tyr Asn Ile Val His Met
 450                 455                 460                 46

GTG GAA TTT CTC ACA ACG CCT TGG GAT TAA GGT GAA GAT AGG ACA ATT   1500
 Trp Asn Phe Ser Gln Arg Leu Gly Ile Lys Val Lys Ile Gly Gln Phe
5               470                 475                 480

CAG CCC ACA TTT TCC ACA GGG TCA ACA GTT ACA CTT ATA TGT AGA CAT   1548
 Ser Pro His Phe Pro Gln Gly Gln Gln Leu His Leu Tyr Val Asp Met
            485                 490                 495

GAC TGA GTT GGC TAC AGG AAG TAG AAA GAT GCC ATC CTC AGT GTG CAG   1596
 Thr Glu Leu Ala Thr Gly Ser Arg Lys Met Pro Ser Ser Val Cys Ser
        500                 505                 510

TGC AGA TTG CCA TCC TGG ATT CAG AAG AAT CTG GAA GGA GGA AAT GGC   1644
 Ala Asp Cys His Pro Gly Phe Arg Arg Ile Trp Lys Glu Glu Met Ala
    515                 520                 525

AGC CTG CTG TTT TGT TTG CAA CCC CTG CCC TGA AAA TGA AAT TTC TAA   1692
 Ala Cys Cys Phe Val Cys Asn Pro Cys Pro Glu Asn Glu Ile Ser Asn
 530                 535                 540                 54

TGA GAC GAT GGT GGT ATT TTG GGT CTT CGT GAA GCA CCA TGA CAC TCC   1740
 Glu Thr Met Val Val Phe Trp Val Phe Val Lys His His Asp Thr Pro
5               550                 555                 560

TAT TGT GAA GGC CAA TAA CAG AAT CCT CAG CTA CCT ATT AAT CGT GTC   1788
 Ile Val Lys Ala Asn Asn Arg Ile Leu Ser Tyr Leu Leu Ile Val Ser
            565                 570                 575

ACT CAT GTT CTG TTT TCT GTG CTC CTT TTT CTT CAT TGG CTA TCC TAA   1836
 Leu Met Phe Cys Phe Leu Cys Ser Phe Phe Phe Ile Gly Tyr Pro Asn
        580                 585                 590

CAG AGC AAC CTG TAT CTT ACA GCA AAT CAC ATT TGG AAT CTT CTT TAC   1884
 Arg Ala Thr Cys Ile Leu Gln Gln Ile Thr Phe Gly Ile Phe Phe Thr
    595                 600                 605

TGT GGC TAT TTC CAC AGT TCT GGC CAA AAC AAT CAC TGT GGT TCT GGC   1932
 Val Ala Ile Ser Thr Val Leu Ala Lys Thr Ile Thr Val Val Leu Ala
 610                 615                 620                 62

TTT CAA AGT CAC AGA CCC AGG AAG ACA ATT AAG AAT CTT TTT GGT ATC   1980
 Phe Lys Val Thr Asp Pro Gly Arg Gln Leu Arg Ile Phe Leu Val Ser
5               630                 635                 640

GGG GAC ACC CAA CTA CAT TAT TCC CAT ATG TTC CCT ATT GCA ATG TAT   2028
 Gly Thr Pro Asn Tyr Ile Ile Pro Ile Cys Ser Leu Leu Gln Cys Ile
            645                 650                 655
```

```
TCT GTG TGC AAT CTG GCT AGC AGT TTC TCC TCC CTT TGT TGA TAT TGA    2076
 Leu Cys Ala Ile Trp Leu Ala Val Ser Pro Pro Phe Val Asp Ile Asp
         660                 665                 670

TGA ACA CTC TGA GCA TGG CCA CAT CAT CAT TGT GTG CAA CAA GGG CTC    2124
 Glu His Ser Glu His Gly His Ile Ile Ile Val Cys Asn Lys Gly Ser
     675                 680                 685

CAT TAC TGC ATT CTA CTG TGT CCT GGG ATA CTT GGC CTG CCT GGC CTT    2172
 Ile Thr Ala Phe Tyr Cys Val Leu Gly Tyr Leu Ala Cys Leu Ala Phe
 690                 695                 700                 70

TGG AAG CTT CAC TAT AGC TTT CTT GGC AAA GAA CCT GCC TGA CAC ATT    2220
 Gly Ser Phe Thr Ile Ala Phe Leu Ala Lys Asn Leu Pro Asp Thr Phe
5                710                 715                 720

CAA CGA AGC CAA GTT CTT GAC CTT CAG CAT GCT AGT GTT CTG CGC TGT    2268
 Asn Glu Ala Lys Phe Leu Thr Phe Ser Met Leu Val Phe Cys Ala Val
             725                 730                 735

CTG GGT CAC CTT CCT CCC TGT CTA CCA TAG CAC CAA GGG CAA GGT CAT    2316
 Trp Val Thr Phe Leu Pro Val Tyr His Ser Thr Lys Gly Lys Val Met
         740                 745                 750

GGT TGC TGT GGA GAT CTT CTC CAT CTT GGC ATC TAG TGC AGG GAT GCT    2364
 Val Ala Val Glu Ile Phe Ser Ile Leu Ala Ser Ser Ala Gly Met Leu
     755                 760                 765

GGG ATG CAT CTT TGC ACC CAA AGT TTA CAT CAT TTT AAT GAG ACC AGA    2412
 Gly Cys Ile Phe Ala Pro Lys Val Tyr Ile Ile Leu Met Arg Pro Asp
 770                 775                 780                 78

CAG AAA TTC GAT CCA CAA AAT CAG GGA GAA ATC ATA TTT C TGAAAAGGTA   2462
 Arg Asn Ser Ile His Lys Ile Arg Glu Lys Ser Tyr Phe
5                790                 795

TTTCAGGAAT TCTGTCAAAT GTAAAGTTGA TACATACACC CCAAATATTT AGTTACAGAG   2522
CATATATCTA GTTTTAGAAT CACTCTCACT GGTTCCTCTA GTTAAGCATA GAAGTACCAT   2582
ATGTACTGAT CTTGCATATG TTGTCTATAA AATCTTACAA TCATTCATTT GCTTAGTATC   2642
TTCTGGAAGA AGTAAAATTT TCAAATAACT AGTACAATTT TATTCATTAT TTTGCTTTCA   2702
TGAGGATTTC CCCCTGGTAA CTTCAAATAA ATTTTATAAG TCAGTTGAAT ATATAACCTT   2762
ACATAGAAAG TGAGTTCTAG GACAGACAGG GATTATACAT AGAAACAAAC TAACTAAAAA   2822
TCAACAAAGA TGAAATCAGA ACACATTTTC TTATTTCCAG TAGGAACACA TACTTGACAG   2882
AATACTGTCT TTTTTTCAGC TGCTCTTTAA GATATTGGCC AATAGTCTAA GCTGAAAATG   2942
TTCTTTATCT ACTCTCAAAT ACAAAAATAT TATATCCAAC AATGGACAGA ATCTGAGAAC   3002
TCCTGTGGTT GAGTTAGGGA ATAGTTGGAA GATACTGAGA AGGAGGTGAC CCATAGGAAT   3062
ACAAAGCAGT CTCAACTAAC CTGGACAACC AAGGTCCCTC AGACACTGAG CCACTAACAA   3122
GTCAGCCTAC TCCAGCTGTT ATGAGGCCCC CAAAACATAT GCAACATAGG ATTGCCTGGT   3182
CCAGCCTCAG CAAGAGAATA CACACCTAAC CACAGAGAGA CTTCCCCAAG GGATTGGGGA   3242
GGTCTGGGGT TTGGAGAGTT GCGGATTGTC CCTTGATGAT TGGAAGGAGG TATTGGATGA   3302
GAATGAATCA GGGGGAAGAC TAGGAAGGGG ATAATGATGG AACTGTAAAA AAAAAAA      3359
```

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 798 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

```
Met Val Ile Phe Phe Leu Leu Asn Ile Pro Phe Leu Leu Ala Asn Phe
 1               5                  10                  15
Met Asp Pro Arg Cys Phe Trp Lys Ile Asn Leu Asn Glu Ile Lys Asp
            20                  25                  30
Glu Val Leu Gly Met Thr Cys Ser Phe Ile Leu Glu Thr Val Gln Lys
        35                  40                  45
Thr Met Asp Lys Asp Tyr Phe Asn Gln Thr Leu Asn Val Leu Asn Thr
    50                  55                  60
Thr Thr Asn His Lys Tyr Ala Leu Ala Leu Ala Phe Thr Val Asp Glu
65                  70                  75                  80
Ile Asn Arg Asn Pro Asp Leu Leu Pro Asn Met Ser Leu Ile Ile Lys
                85                  90                  95
Tyr Asn Leu Gly His Cys Asp Gly Lys Thr Val Thr Thr Leu Ser Asp
            100                 105                 110
Leu Phe Asn Pro Asn Asn His Leu His Phe Pro Asn Tyr Leu Cys Asn
        115                 120                 125
Glu Gly Ile Met Cys Leu Val Leu Leu Thr Gly Pro His Trp Arg Ala
    130                 135                 140
Ser Leu Tyr Leu Trp Ile Ser Val Tyr Val Tyr Leu Ser Pro His Phe
145                 150                 155                 160
Leu Gln Leu Ser Tyr Gly Pro Phe Tyr Ser Ile Phe Ser Asp Asn Glu
                165                 170                 175
Gln Tyr Pro Tyr Leu Tyr Gln Met Gly Pro Lys Asp Ser Ser Leu Ala
            180                 185                 190
Leu Ala Met Val Ser Phe Ile Ile Tyr Phe Lys Trp Asn Trp Val Gly
        195                 200                 205
Leu Phe Ile Ser Asp Asp Asp Gln Gly Asn Gln Phe Leu Ser Glu Leu
    210                 215                 220
Lys Lys Glu Ser Gln Thr Lys Asp Ile Cys Phe Ala Phe Val Asn Met
225                 230                 235                 240
Ile Ser Val Ser Asp Val Ser Tyr Tyr His Lys Thr Glu Met Tyr Tyr
                245                 250                 255
Asn Gln Ile Val Met Ser Ser Thr Lys Val Ile Ile Ile Tyr Gly Glu
            260                 265                 270
Thr Asn Ser Ile Ile Glu Leu Ser Phe Arg Met Trp Ser Ser Pro Val
        275                 280                 285
Lys Gln Arg Ile Trp Val Thr Thr Lys Gln Phe Asp Cys Pro Thr Ser
    290                 295                 300
Lys Arg Asp Leu Thr His Gly Thr Phe Tyr Gly Thr Leu Thr Phe Leu
305                 310                 315                 320
His His Tyr Gly Glu Ile Ser Gly Phe Lys Asn Phe Val Gln Thr Arg
                325                 330                 335
Tyr Asn Leu Arg Ser Thr Asp Leu Tyr Leu Val Met Pro Glu Trp Lys
            340                 345                 350
Tyr Phe Asn Tyr Glu Ala Ser Ala Ser Asn Cys Lys Ile Leu Arg Asn
        355                 360                 365
Tyr Leu Ser Asn Ile Ser Leu Glu Trp Leu Met Glu Gln Lys Phe Asp
    370                 375                 380
Met Ser Phe Ser Asp Tyr Ser His Asn Ile Tyr Asn Ala Val Tyr Ala
385                 390                 395                 400
Ile Ala His Ala Leu His Glu Lys Asn Leu Gln Glu Val Glu Asn Gln
                405                 410                 415
Ala Ile Asn Asn Ala Lys Gly Glu Asn Thr His Cys Leu Lys Leu Asn
            420                 425                 430
Ser Phe Leu Arg Lys Thr His Phe Thr Asn Ser Leu Gly Asn Arg Val
        435                 440                 445
Ile Met Lys Gln Arg Glu Val Val His Gly Asp Tyr Asn Ile Val His
    450                 455                 460
Met Trp Asn Phe Ser Gln Arg Leu Gly Ile Lys Val Lys Ile Gly Gln
465                 470                 475                 480
Phe Ser Pro His Phe Pro Gln Gly Gln Gln Leu His Leu Tyr Val Asp
                485                 490                 495
Met Thr Glu Leu Ala Thr Gly Ser Arg Lys Met Pro Ser Ser Val Cys
            500                 505                 510
Ser Ala Asp Cys His Pro Gly Phe Arg Arg Ile Trp Lys Glu Glu Met
```

```
        515                 520                 525
Ala Ala Cys Cys Phe Val Cys Asn Pro Cys Pro Glu Asn Glu Ile Ser
    530                 535                 540
Asn Glu Thr Met Val Val Phe Trp Val Phe Val Lys His His Asp Thr
545                 550                 555                 560
Pro Ile Val Lys Ala Asn Asn Arg Ile Leu Ser Tyr Leu Leu Ile Val
                565                 570                 575
Ser Leu Met Phe Cys Phe Leu Cys Ser Phe Phe Phe Ile Gly Tyr Pro
            580                 585                 590
Asn Arg Ala Thr Cys Ile Leu Gln Gln Ile Thr Phe Gly Ile Phe Phe
        595                 600                 605
Thr Val Ala Ile Ser Thr Val Leu Ala Lys Thr Ile Thr Val Val Leu
    610                 615                 620
Ala Phe Lys Val Thr Asp Pro Gly Arg Gln Leu Arg Ile Phe Leu Val
625                 630                 635                 640
Ser Gly Thr Pro Asn Tyr Ile Ile Pro Ile Cys Ser Leu Leu Gln Cys
                645                 650                 655
Ile Leu Cys Ala Ile Trp Leu Ala Val Ser Pro Pro Phe Val Asp Ile
            660                 665                 670
Asp Glu His Ser Glu His Gly His Ile Ile Ile Val Cys Asn Lys Gly
        675                 680                 685
Ser Ile Thr Ala Phe Tyr Cys Val Leu Gly Tyr Leu Ala Cys Leu Ala
    690                 695                 700
Phe Gly Ser Phe Thr Ile Ala Phe Leu Ala Lys Asn Leu Pro Asp Thr
705                 710                 715                 720
Phe Asn Glu Ala Lys Phe Leu Thr Phe Ser Met Leu Val Phe Cys Ala
                725                 730                 735
Val Trp Val Thr Phe Leu Pro Val Tyr His Ser Thr Lys Gly Lys Val
            740                 745                 750
Met Val Ala Val Glu Ile Phe Ser Ile Leu Ala Ser Ser Ala Gly Met
        755                 760                 765
Leu Gly Cys Ile Phe Ala Pro Lys Val Tyr Ile Ile Leu Met Arg Pro
    770                 775                 780
Asp Arg Asn Ser Ile His Lys Ile Arg Glu Lys Ser Tyr Phe
785                 790                 795
```

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3012 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 3...2087
(D) OTHER INFORMATION: GoVN13B

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

```
AT GTC TAC CTG TCT CCA CAT TTC CTT CAG CTT TCC TAT GGA CCT TTC      47
   Val Tyr Leu Ser Pro His Phe Leu Gln Leu Ser Tyr Gly Pro Phe
    1               5                  10                  15

TAC TCC ATC TTC AGT GAT AAT GAA CAA TAT CCT TAT CTC TAT CAG ATG      95
Tyr Ser Ile Phe Ser Asp Asn Glu Gln Tyr Pro Tyr Leu Tyr Gln Met
                20                  25                  30

GGC CCA AAG GAC TCA TCA CTA GCA TTG GCA ATG GTC TCC TTC ATA ATT     143
Gly Pro Lys Asp Ser Ser Leu Ala Leu Ala Met Val Ser Phe Ile Ile
            35                  40                  45
```

```
TAC TTC AAG TGG AAC TGG GTT GGG CTA TTT ATC TCA GAT GAT GAT CAA      191
Tyr Phe Lys Trp Asn Trp Val Gly Leu Phe Ile Ser Asp Asp Asp Gln
        50                  55                  60

GGC AAT CAA TTT CTC TCA GAG TTG AAA AAA GAG AGC CAA ACC AAG GAT      239
Gly Asn Gln Phe Leu Ser Glu Leu Lys Lys Glu Ser Gln Thr Lys Asp
    65                  70                  75

ATT TGC TTT GCC TTT GTG AAC ATG ATA TCA GTC AGT GAT GTT TCA TAC      287
Ile Cys Phe Ala Phe Val Asn Met Ile Ser Val Ser Asp Val Ser Tyr
80                  85                  90                  95

TAT CAT AAA ACT GAA ATG TAC TAC AAC CAA ATT GTG ATG TCA TCC ACA      335
Tyr His Lys Thr Glu Met Tyr Tyr Asn Gln Ile Val Met Ser Ser Thr
                100                 105                 110

AAG GTT ATT ATC ATT TAT GGG GAA ACA AAC AGT ATT ATT GAA TTG AGC      383
Lys Val Ile Ile Ile Tyr Gly Glu Thr Asn Ser Ile Ile Glu Leu Ser
            115                 120                 125

TTC AGA ATG TGG TCA TCT CCA GTT AAA CAG AGA ATA TGG GTC ACC ACA      431
Phe Arg Met Trp Ser Ser Pro Val Lys Gln Arg Ile Trp Val Thr Thr
        130                 135                 140

AAA CAA TTT GAT TGC CCT ACC AGT AAG AGA GAC TTA ACT CAT GGC ACA      479
Lys Gln Phe Asp Cys Pro Thr Ser Lys Arg Asp Leu Thr His Gly Thr
    145                 150                 155

TTC TAT GGG ACC CTT ACA TTT CTA CAC CAC TAT GGT GAG ATT TCT GGC      527
Phe Tyr Gly Thr Leu Thr Phe Leu His His Tyr Gly Glu Ile Ser Gly
160                 165                 170                 175

TTT AAA AAT TTT GTA CAG ACA CGG TAC AAT CTC AGA AGC ACA GAT TTA      575
Phe Lys Asn Phe Val Gln Thr Arg Tyr Asn Leu Arg Ser Thr Asp Leu
                180                 185                 190

TAT CTA GTA ATG CCA GAG TGG AAA TAT TTT AAC TAT GAA GCC TCA GCA      623
Tyr Leu Val Met Pro Glu Trp Lys Tyr Phe Asn Tyr Glu Ala Ser Ala
            195                 200                 205

TCT AAC TGT AAA ATA CTG AGA AAC TAT TTA TCC AAT ATC TCA CTG GAA      671
Ser Asn Cys Lys Ile Leu Arg Asn Tyr Leu Ser Asn Ile Ser Leu Glu
        210                 215                 220

TGG CTA ATG GAA CAG AAA TTT GAC ATG TCA TTT AGT GAT TAT AGT CAC      719
Trp Leu Met Glu Gln Lys Phe Asp Met Ser Phe Ser Asp Tyr Ser His
    225                 230                 235

AAC ATA TAC AAT GCT GTA TAT GCC ATT GCT CAT GCA CTC CAT GAG AAA      767
Asn Ile Tyr Asn Ala Val Tyr Ala Ile Ala His Ala Leu His Glu Lys
240                 245                 250                 255

GAT CTG CAA GAA TTT GAA AAT CAG GCA ATA AAC AAT GCG AAA GGA GAA      815
Asp Leu Gln Glu Phe Glu Asn Gln Ala Ile Asn Asn Ala Lys Gly Glu
                260                 265                 270

AAT ACT CAC TGC TTG AAG CTA AAC TCA TTT CTG AGA AAG ACC CAC TTC      863
Asn Thr His Cys Leu Lys Leu Asn Ser Phe Leu Arg Lys Thr His Phe
            275                 280                 285

ACT AAT TCT CTT GGG AAC AGA GTA ATT ATG AAA CAG AGA GAA GTA GTG      911
Thr Asn Ser Leu Gly Asn Arg Val Ile Met Lys Gln Arg Glu Val Val
        290                 295                 300

CAT GGA GAC TAT AAT ATT GTT CAC ATG TGG AAT TTC TCA CAA CGC CTT      959
```

```
His Gly Asp Tyr Asn Ile Val His Met Trp Asn Phe Ser Gln Arg Leu
    305                 310                 315

GGG ATT AAG GTG AAG ATA GGA CAA TTC AGC CCA CAT TTT CCA CAG GGT    1007
Gly Ile Lys Val Lys Ile Gly Gln Phe Ser Pro His Phe Pro Gln Gly
320                 325                 330                 335

CAA CAG TTA CAC TTA TAT GTA GAC ATG ACT GAG TTG GCT ACA GGA AGT    1055
Gln Gln Leu His Leu Tyr Val Asp Met Thr Glu Leu Ala Thr Gly Ser
                340                 345                 350

AGA AAG ATG CCA TCC TCA GTG TGC AGT GCA GAT TGC CAT CCT GGA TTC    1103
Arg Lys Met Pro Ser Ser Val Cys Ser Ala Asp Cys His Pro Gly Phe
            355                 360                 365

AGA AGA ATC TGG AAG GAG GAA ATG GCA GCC TGC TGT TTT GTT TGC AAC    1151
Arg Arg Ile Trp Lys Glu Glu Met Ala Ala Cys Cys Phe Val Cys Asn
        370                 375                 380

CCC TGC CCT GAA AAT GAA ATT TCT AAT GAG ACG AAT ATG GAT CAG TGT    1199
Pro Cys Pro Glu Asn Glu Ile Ser Asn Glu Thr Asn Met Asp Gln Cys
    385                 390                 395

GCG AAT TGT CCA GAA TAC CAG TAT GCC AAC ACA GAA AAG AAC AAA TGC    1247
Ala Asn Cys Pro Glu Tyr Gln Tyr Ala Asn Thr Glu Lys Asn Lys Cys
400                 405                 410                 415

ATC CAG AAA GGT GTG ATT GTT CTA AGC TAT GAA GAC CCC TTG GGG ATG    1295
Ile Gln Lys Gly Val Ile Val Leu Ser Tyr Glu Asp Pro Leu Gly Met
                420                 425                 430

GCT CTT GCC TTA ATA GCA TTC TGT TTC TCT GCA TTC ACA GTG GTG GTA    1343
Ala Leu Ala Leu Ile Ala Phe Cys Phe Ser Ala Phe Thr Val Val Val
            435                 440                 445

TTT TGG GTC TTC GTG AAG CAC CAT GAC ACT CCT ATT GTG AAG GCC AAT    1391
Phe Trp Val Phe Val Lys His His Asp Thr Pro Ile Val Lys Ala Asn
        450                 455                 460

AAC AGA ATC CTC AGC TAC CTA TTA ATC GTG TCA CTC ATG TTC TGT TTT    1439
Asn Arg Ile Leu Ser Tyr Leu Leu Ile Val Ser Leu Met Phe Cys Phe
    465                 470                 475

CTG TGC TCC TTT TTC TTC ATT GGC TAT CCT AAC AGA GCA ACC TGT ATC    1487
Leu Cys Ser Phe Phe Phe Ile Gly Tyr Pro Asn Arg Ala Thr Cys Ile
480                 485                 490                 495

TTA CAG CAA ATC ACA TTT GGA ATC TTC TTT ACT GTG GCT ATT TCC ACA    1535
Leu Gln Gln Ile Thr Phe Gly Ile Phe Phe Thr Val Ala Ile Ser Thr
                500                 505                 510

GTT CTG GCC AAA ACA ATC ACT GTG GTT CTG GCT TTC AAA GTC ACA GAC    1583
Val Leu Ala Lys Thr Ile Thr Val Val Leu Ala Phe Lys Val Thr Asp
            515                 520                 525

CCA GGA AGA CAA TTA AGA ATC TTT TTG GTA TCG GGG ACA CCC AAC TAC    1631
Pro Gly Arg Gln Leu Arg Ile Phe Leu Val Ser Gly Thr Pro Asn Tyr
        530                 535                 540

ATT ATT CCC ATA TGT TCC CTA TTG CAA TGT ATT CTG TGT GCA ATC TGG    1679
Ile Ile Pro Ile Cys Ser Leu Leu Gln Cys Ile Leu Cys Ala Ile Trp
    545                 550                 555

CTA GCA GTT TCT CCT CCC TTT GTT GAT ATT GAT GAA CAC TCT GAG CAT    1727
Leu Ala Val Ser Pro Pro Phe Val Asp Ile Asp Glu His Ser Glu His
```

```
560                 565                 570                 575

GGC CAC ATC ATC ATT GTG TGC AAC AAG GGC TCC ATT ACT GCA TTC TAC    1775
Gly His Ile Ile Ile Val Cys Asn Lys Gly Ser Ile Thr Ala Phe Tyr
                580                 585                 590

TGT GTC CTG GGA TAC TTG GCC TGC CTG GCC TTT GGA AGC TTC ACT ATA    1823
Cys Val Leu Gly Tyr Leu Ala Cys Leu Ala Phe Gly Ser Phe Thr Ile
            595                 600                 605

GCT TTC TTG GCA AAG AAC CTG CCT GAC ACA TTC AAC GAA GCC AAG TTC    1871
Ala Phe Leu Ala Lys Asn Leu Pro Asp Thr Phe Asn Glu Ala Lys Phe
        610                 615                 620

TTG ACC TTC AGC ATG CTA GTG TTC TGC GCT GTC TGG GTC ACC TTC CTC    1919
Leu Thr Phe Ser Met Leu Val Phe Cys Ala Val Trp Val Thr Phe Leu
    625                 630                 635

CCT GTC TAC CAT AGC ACC AAG GGC AAG GTC ATG GTT GCT GTG GAG ATC    1967
Pro Val Tyr His Ser Thr Lys Gly Lys Val Met Val Ala Val Glu Ile
640                 645                 650                 655

TTC TCC ATC TTG GCA TCT AGT GCA GGG ATG CTG GGA TGC ATC TTT GCA    2015
Phe Ser Ile Leu Ala Ser Ser Ala Gly Met Leu Gly Cys Ile Phe Ala
                660                 665                 670

CCC AAA GTT TAC ATC ATT TTA ATG AGA CCA GAC AGA AAT TCG ATC CAC    2063
Pro Lys Val Tyr Ile Ile Leu Met Arg Pro Asp Arg Asn Ser Ile His
            675                 680                 685

AAA ATC AGG GAG AAA TCA TAT TTC TGAAAAGGTA TTTCAGGAAT TCTGTCAAAT   2117
Lys Ile Arg Glu Lys Ser Tyr Phe
        690                 695

GTAAAGTTGA TACATACACC CCAAATATTT AGTTACAGAG CATATATCTA GTTTTAGAAT   2177
CACTCTCACT GGTTCCTCTA GTTATGCATA GAAGTACCAT ATGTACTGAT CTTGCATATG   2237
TTGTCTATAA AATCTTACAA TCATTCATTT GCTTAGTATC TTCTGGAAGA AGTAAAATTT   2297
TCAAATAACT AGTACAATTT TATTCATTAT TTTGCTTTCA TGAGGATTTC CCCCTGGTAA   2357
CTTCAAATAA ATTTTATAAG TCAGTTGAAT ATATAACCTT ACATAGAAAG TGAGTTCTAG   2417
GACAGACAGG GATTATACAT AGAAACAAAC TAACTAAAAA TCAACAAAGA TGAAATCAGA   2477
ACACATTTTC TTATTTCCAG TAGGAACACA TACTTGACAG AATACTGTCT TTTTTTCAGC   2537
TGCTCTTTAA GATATTGGCC AATAGTCTAA GCTGAAAATG TTCTTTATCT ACTCTCAAAT   2597
ACAAAAATAT TATATCCAAC AATGGACAGA ATCTGAGAAC TCCTGTGGTT GAGTTAGGGA   2657
ATAGTTGGAA GATACTGAGA AGGAGGGTGA CCCATAGGAA TACAAAGCAG TCTCAACTAA   2717
CCTGGACAAC CAAGGTCCCT CAGACACTGA GCCACTAACA AGTCAGCCTA CTCCAGCTGT   2777
TATGAGGCCC CCAAAACATA TGCAACATAG GATTGCCTGG TCCAGCCTCA GCAAGAGAAT   2837
ACACACCTAA CCACAGAGAG ACTTCCCCAA GGGATTGGGG AGGTCTGGGG TTTGGAGAGT   2897
TGCGGATTGT CCCTTGATGA TTGGAAGGAG GTATTGGATG AGAATGAATC AGGGGGAAGA   2957
CTAGGAAGGG GATAATGATG GAACTGTAAA AAAAATTAAA AAAAAAAAAA AAAAA        3012
```

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 695 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

```
Val Tyr Leu Ser Pro His Phe Leu Gln Leu Ser Tyr Gly Pro Phe Tyr
 1               5                  10                  15
```

Ser Ile Phe Ser Asp Asn Glu Gln Tyr Pro Tyr Leu Tyr Gln Met Gly
20 25 30
Pro Lys Asp Ser Ser Leu Ala Leu Ala Met Val Ser Phe Ile Ile Tyr
35 40 45
Phe Lys Trp Asn Trp Val Gly Leu Phe Ile Ser Asp Asp Asp Gln Gly
50 55 60
Asn Gln Phe Leu Ser Glu Leu Lys Lys Glu Ser Gln Thr Lys Asp Ile
65 70 75 80
Cys Phe Ala Phe Val Asn Met Ile Ser Val Ser Asp Val Ser Tyr Tyr
85 90 95
His Lys Thr Glu Met Tyr Tyr Asn Gln Ile Val Met Ser Ser Thr Lys
100 105 110
Val Ile Ile Ile Tyr Gly Glu Thr Asn Ser Ile Ile Glu Leu Ser Phe
115 120 125
Arg Met Trp Ser Ser Pro Val Lys Gln Arg Ile Trp Val Thr Thr Lys
130 135 140
Gln Phe Asp Cys Pro Thr Ser Lys Arg Asp Leu Thr His Gly Thr Phe
145 150 155 160
Tyr Gly Thr Leu Thr Phe Leu His His Tyr Gly Glu Ile Ser Gly Phe
165 170 175
Lys Asn Phe Val Gln Thr Arg Tyr Asn Leu Arg Ser Thr Asp Leu Tyr
180 185 190
Leu Val Met Pro Glu Trp Lys Tyr Phe Asn Tyr Glu Ala Ser Ala Ser
195 200 205
Asn Cys Lys Ile Leu Arg Asn Tyr Leu Ser Asn Ile Ser Leu Glu Trp
210 215 220
Leu Met Glu Gln Lys Phe Asp Met Ser Phe Ser Asp Tyr Ser His Asn
225 230 235 240
Ile Tyr Asn Ala Val Tyr Ala Ile Ala His Ala Leu His Glu Lys Asp
245 250 255
Leu Gln Glu Phe Glu Asn Gln Ala Ile Asn Asn Ala Lys Gly Glu Asn
260 265 270
Thr His Cys Leu Lys Leu Asn Ser Phe Leu Arg Lys Thr His Phe Thr
275 280 285
Asn Ser Leu Gly Asn Arg Val Ile Met Lys Gln Arg Glu Val Val His
290 295 300
Gly Asp Tyr Asn Ile Val His Met Trp Asn Phe Ser Gln Arg Leu Gly
305 310 315 320
Ile Lys Val Lys Ile Gly Gln Phe Ser Pro His Phe Pro Gln Gly Gln
325 330 335
Gln Leu His Leu Tyr Val Asp Met Thr Glu Leu Ala Thr Gly Ser Arg
340 345 350
Lys Met Pro Ser Ser Val Cys Ser Ala Asp Cys His Pro Gly Phe Arg
355 360 365
Arg Ile Trp Lys Glu Glu Met Ala Ala Cys Cys Phe Val Cys Asn Pro
370 375 380
Cys Pro Glu Asn Glu Ile Ser Asn Glu Thr Asn Met Asp Gln Cys Ala
385 390 395 400
Asn Cys Pro Glu Tyr Gln Tyr Ala Asn Thr Glu Lys Asn Lys Cys Ile
405 410 415
Gln Lys Gly Val Ile Val Leu Ser Tyr Glu Asp Pro Leu Gly Met Ala
420 425 430
Leu Ala Leu Ile Ala Phe Cys Phe Ser Ala Phe Thr Val Val Val Phe
435 440 445
Trp Val Phe Val Lys His His Asp Thr Pro Ile Val Lys Ala Asn Asn
450 455 460
Arg Ile Leu Ser Tyr Leu Leu Ile Val Ser Leu Met Phe Cys Phe Leu
465 470 475 480
Cys Ser Phe Phe Phe Ile Gly Tyr Pro Asn Arg Ala Thr Cys Ile Leu
485 490 495
Gln Gln Ile Thr Phe Gly Ile Phe Phe Thr Val Ala Ile Ser Thr Val
500 505 510
Leu Ala Lys Thr Ile Thr Val Val Leu Ala Phe Lys Val Thr Asp Pro
515 520 525
Gly Arg Gln Leu Arg Ile Phe Leu Val Ser Gly Thr Pro Asn Tyr Ile

```
    530                 535                 540
Ile Pro Ile Cys Ser Leu Leu Gln Cys Ile Leu Cys Ala Ile Trp Leu
545                 550                 555                 560
Ala Val Ser Pro Pro Phe Val Asp Ile Asp Glu His Ser Glu His Gly
                565                 570                 575
His Ile Ile Ile Val Cys Asn Lys Gly Ser Ile Thr Ala Phe Tyr Cys
            580                 585                 590
Val Leu Gly Tyr Leu Ala Cys Leu Ala Phe Gly Ser Phe Thr Ile Ala
        595                 600                 605
Phe Leu Ala Lys Asn Leu Pro Asp Thr Phe Asn Glu Ala Lys Phe Leu
    610                 615                 620
Thr Phe Ser Met Leu Val Phe Cys Ala Val Trp Val Thr Phe Leu Pro
625                 630                 635                 640
Val Tyr His Ser Thr Lys Gly Lys Val Met Val Ala Val Glu Ile Phe
                645                 650                 655
Ser Ile Leu Ala Ser Ser Ala Gly Met Leu Gly Cys Ile Phe Ala Pro
            660                 665                 670
Lys Val Tyr Ile Ile Leu Met Arg Pro Asp Arg Asn Ser Ile His Lys
        675                 680                 685
Ile Arg Glu Lys Ser Tyr Phe
    690                 695
```

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 435 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

```
CAGACTCTGA GCTACACCCT CCTTGTCTCC CTCACACTCT GCTTTCTCTC TTCCTCGCTC     60
TTCATCGGCC GCCCCAGCCC TGCCACCTGC CTCCTCTCAC AGACCACCTT TGCAGCTGTG    120
TTCACAGTGG CTGTGTTTTT CTGCAGGGCC TTCCAGGCTA TAAGGCCAGA AAGCAGGATC    180
CGAAAGTGGA TGGGTCCCCA AAAAACAAAT TCTGTTGTCT TCCTTTGCTC CTTTACCCAA    240
GTGACCCTCT GTGGAATCTG GCTGGGGACA GAGCCTCCCT TCGTAAACAA GGACCCTCAG    300
TTCATGCCTG GCTACATCAT TATCCAGTGT AATGAGGGCT CCGTCACTGC CTTCTACTCT    360
GTCTTGGGCT ACTTGGGCTT CTTGGTTTTA GGGTCCCTTG CTGTAGCCTT TCTGGCAAGG    420
AACCTGCCTG ATGCT                                                     435
```

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 145 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

```
Gln Thr Leu Ser Tyr Thr Leu Leu Val Ser Leu Thr Leu Cys Phe Leu
 1               5                  10                  15
Ser Ser Ser Leu Phe Ile Gly Arg Pro Ser Pro Ala Thr Cys Leu Leu
            20                  25                  30
Ser Gln Thr Thr Phe Ala Ala Val Phe Thr Val Ala Val Phe Phe Cys
        35                  40                  45
Arg Ala Phe Gln Ala Ile Arg Pro Glu Ser Arg Ile Arg Lys Trp Met
    50                  55                  60
Gly Pro Gln Lys Thr Asn Ser Val Val Phe Leu Cys Ser Phe Thr Gln
65                  70                  75                  80
```

```
Val Thr Leu Cys Gly Ile Trp Leu Gly Thr Glu Pro Pro Phe Val Asn
                85                  90                  95
Lys Asp Pro Gln Phe Met Pro Gly Tyr Ile Ile Ile Gln Cys Asn Glu
            100                 105                 110
Gly Ser Val Thr Ala Phe Tyr Ser Val Leu Gly Tyr Leu Gly Phe Leu
        115                 120                 125
Val Leu Gly Ser Leu Ala Val Ala Phe Leu Ala Arg Asn Leu Pro Asp
    130                 135                 140
Ala
145
```

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 474 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

```
CCCATTGTGA AGGCTAATAA CCAGACTCTG AGCTACACCC TCCTTGTCTC CCTCACACTC     60
TGCTTTCTCT CTTCCTCGCT CTTCATCGGC CGCCCCAGCC CTGCCACCTG CCTCCTCTCA    120
CAGACCACCT TTGCAGCTGT GTTCACAGTG GCTGTGTTTT CTGCAGGGCC TTCCAGGCTA    180
TAAGGCCAGA AAGCAGGATC CGAAAGTGGA TGGGTCCCCA AAAAACAAAT TCTGTTGTCT    240
TCCTTTGCTC CTTTACCCAA GTGACCCTCT GTGGAATCTG GCTGGGGACA GAGCCTCCCT    300
TCGTAAACAA GGACCCTCAG TTCATGCCTG GCTACATCAT TATCCAGTGT AATGAGGGCT    360
CCGTCACTGC CTTCTACTCT GTCTTGGGCT ACTTGGGCTT CTTGGTTTTA GGGTCCCTTG    420
CTGTAGCCTT TCTGGCAAGG AACCCGCCAG ATACGTTCAA TGAGGCCAAG TTAA          474
```

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 338 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

```
ACTCCCATTG TGAAGGCCAA CAACTGCCAG CTCAGCTATC TCCTGCTGTC CTCCTTGGCC     60
CTCAGCTTCC TCTGCCCCTT CATGTTCATT GGCCACCCAG ACCCCATCAC TTGTGCTGTG    120
CACNAGGCAG ATTTTGGGGT CACCTTCATG GTCTGCACAT CCACTGTGCT GGCCAAGACC    180
ATCGTGGTGG TGGCAGCCTT CCATGCCACC CAGGCAGACA CTCAGCTTAG GGGGTGGGCG    240
GGGACAGTCC TCCTCAGCAC CATCCTCACT GTTCCCTGAC CCAGGCAGCC TTGTGTGCAC    300
TCTGGGTGAC CAGATGGCCC CCTCAGCCTG TAAAATCT                            338
```

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 182 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

```
AACCTNCCCG ATACNTTCAA TGAAGCCAAG TTCTTGATGT TCAGCATGCT GATGTTATGT     60
ACTGTTTGAA TTACCTTCCA TACTGTGTAA CATAGCACCA AAGGGAAGGT CATGGTTGCC    120
TTGGAAATAT TCTCCACCTT GACTTCCAGT GCTGAGTGCT AGGNTGTATC TTCGCNCCAA    180
```

AA 182

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

ATTGGATCCA GGCCGCTCTG GACAAAATAT GAATTCT 37

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

GGCACATGGA CGAAATCTTG GTACTCTTCA GAATTCT 37

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 51 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

```
Asn Met Asp Gln Cys Ala Asn Cys Pro Glu Tyr Gln Tyr Ala Asn Thr
 1               5                  10                  15
Glu Lys Asn Lys Cys Ile Gln Lys Gly Val Ile Val Leu Ser Tyr Glu
            20                  25                  30
Asp Pro Leu Gly Met Ala Leu Ala Leu Ile Ala Phe Cys Phe Ser Ala
        35                  40                  45
Phe Thr Val
    50
```

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1079 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

```
Met Ala Ser Tyr Ser Cys Cys Leu Ala Leu Leu Ala Leu Ala Trp His
 1               5                  10                  15
Ser Ser Ala Tyr Gly Pro Asp Gln Arg Ala Gln Lys Lys Gly Asp Ile
```

```
                20                  25                  30
Ile Leu Gly Gly Leu Phe Pro Ile His Phe Gly Val Ala Ala Lys Asp
            35                  40                  45
Gln Asp Leu Lys Ser Arg Pro Glu Ser Val Glu Cys Ile Arg Tyr Asn
        50                  55                  60
Phe Arg Gly Phe Arg Trp Leu Gln Ala Met Ile Phe Ala Ile Glu Glu
65                  70                  75                  80
Ile Asn Ser Ser Pro Ser Leu Leu Pro Asn Met Thr Leu Gly Tyr Arg
                85                  90                  95
Ile Phe Asp Thr Cys Asn Thr Val Ser Lys Ala Leu Glu Ala Thr Leu
            100                 105                 110
Ser Phe Val Ala Gln Asn Lys Ile Asp Ser Leu Asn Leu Asp Glu Phe
        115                 120                 125
Cys Asn Cys Ser Glu His Ile Pro Ser Thr Ile Ala Val Val Gly Ala
    130                 135                 140
Thr Gly Ser Gly Val Ser Thr Ala Val Ala Asn Leu Leu Gly Leu Phe
145                 150                 155                 160
Tyr Ile Pro Gln Val Ser Tyr Ala Ser Ser Ser Arg Leu Leu Ser Asn
                165                 170                 175
Lys Asn Gln Tyr Lys Ser Phe Leu Arg Thr Ile Pro Asn Asp Glu His
            180                 185                 190
Gln Ala Thr Ala Met Ala Asp Ile Ile Glu Tyr Phe Arg Trp Asn Trp
        195                 200                 205
Val Gly Thr Ile Ala Ala Asp Asp Asp Tyr Gly Arg Pro Gly Ile Glu
    210                 215                 220
Lys Phe Arg Glu Glu Ala Glu Glu Arg Asp Ile Cys Ile Asp Phe Ser
225                 230                 235                 240
Glu Leu Ile Ser Gln Tyr Ser Asp Glu Glu Glu Ile Gln Gln Val Val
                245                 250                 255
Glu Val Ile Gln Asn Ser Thr Ala Lys Val Ile Val Val Phe Ser Ser
            260                 265                 270
Gly Pro Asp Leu Glu Pro Leu Ile Lys Glu Ile Val Arg Arg Asn Ile
        275                 280                 285
Thr Gly Arg Ile Trp Leu Ala Ser Glu Ala Trp Ala Ser Ser Ser Leu
    290                 295                 300
Ile Ala Met Pro Glu Tyr Phe His Val Val Gly Gly Thr Ile Gly Phe
305                 310                 315                 320
Gly Leu Lys Ala Gly Gln Ile Pro Gly Phe Arg Glu Phe Leu Gln Lys
                325                 330                 335
Val His Pro Arg Lys Ser Val His Asn Gly Phe Ala Lys Glu Phe Trp
            340                 345                 350
Glu Glu Thr Phe Asn Cys His Leu Gln Glu Gly Ala Lys Gly Pro Leu
        355                 360                 365
Pro Val Asp Thr Phe Val Arg Ser His Glu Glu Gly Gly Asn Arg Leu
    370                 375                 380
Leu Asn Ser Ser Thr Ala Phe Arg Pro Leu Cys Thr Gly Asp Glu Asn
385                 390                 395                 400
Ile Asn Ser Val Glu Thr Pro Tyr Met Asp Tyr Glu His Leu Arg Ile
                405                 410                 415
Ser Tyr Asn Val Tyr Leu Ala Val Tyr Ser Ile Ala His Ala Leu Gln
            420                 425                 430
Asp Ile Tyr Thr Cys Leu Pro Gly Arg Gly Leu Phe Thr Asn Gly Ser
        435                 440                 445
Cys Ala Asp Ile Lys Lys Val Glu Ala Trp Gln Val Leu Lys His Leu
    450                 455                 460
Arg His Leu Asn Phe Thr Asn Asn Met Gly Glu Gln Val Thr Phe Asp
465                 470                 475                 480
Glu Cys Gly Asp Leu Val Gly Asn Tyr Ser Ile Ile Asn Trp His Leu
                485                 490                 495
Ser Pro Glu Asp Gly Ser Ile Val Phe Lys Glu Val Gly Tyr Tyr Asn
            500                 505                 510
Val Tyr Ala Lys Lys Gly Glu Arg Leu Phe Ile Asn Glu Glu Lys Ile
        515                 520                 525
Leu Trp Ser Gly Phe Ser Arg Glu Val Pro Phe Ser Asn Cys Ser Arg
    530                 535                 540
```

```
Asp Cys Gln Ala Gly Thr Arg Lys Gly Ile Ile Glu Gly Glu Pro Thr
545                 550                 555                 560
Cys Cys Phe Glu Cys Val Glu Cys Pro Asp Gly Glu Tyr Ser Gly Glu
                565                 570                 575
Thr Asp Ala Ser Ala Cys Asp Lys Cys Pro Asp Asp Phe Trp Ser Asn
            580                 585                 590
Glu Asn His Thr Ser Cys Ile Ala Lys Glu Ile Glu Phe Leu Ala Trp
        595                 600                 605
Thr Glu Pro Phe Gly Ile Ala Leu Thr Leu Phe Ala Val Leu Gly Ile
    610                 615                 620
Phe Leu Thr Ala Phe Val Leu Gly Val Phe Ile Lys Phe Arg Asn Thr
625                 630                 635                 640
Pro Ile Val Lys Ala Thr Asn Arg Glu Leu Ser Tyr Leu Leu Leu Phe
                645                 650                 655
Ser Leu Leu Cys Cys Phe Ser Ser Ser Leu Phe Phe Ile Gly Glu Pro
            660                 665                 670
Gln Asp Trp Thr Cys Arg Leu Arg Gln Pro Ala Phe Gly Ile Ser Phe
        675                 680                 685
Val Leu Cys Ile Ser Cys Ile Leu Val Lys Thr Asn Arg Val Leu Leu
    690                 695                 700
Val Phe Glu Ala Lys Ile Pro Thr Ser Phe His Arg Lys Trp Trp Gly
705                 710                 715                 720
Leu Asn Leu Gln Phe Leu Leu Val Phe Leu Cys Thr Phe Met Gln Ile
                725                 730                 735
Leu Ile Cys Ile Ile Trp Leu Tyr Thr Ala Pro Pro Ser Ser Tyr Arg
            740                 745                 750
Asn His Glu Leu Glu Asp Glu Ile Ile Phe Ile Thr Cys His Glu Gly
        755                 760                 765
Ser Leu Met Ala Leu Gly Ser Leu Ile Gly Tyr Thr Cys Leu Leu Ala
    770                 775                 780
Ala Ile Cys Phe Phe Phe Ala Phe Lys Ser Arg Lys Leu Pro Glu Asn
785                 790                 795                 800
Phe Asn Glu Ala Lys Phe Ile Thr Phe Ser Met Leu Ile Phe Phe Ile
                805                 810                 815
Val Trp Ile Ser Phe Ile Pro Ala Tyr Ala Ser Thr Tyr Gly Lys Phe
            820                 825                 830
Val Ser Ala Val Glu Val Ile Ala Ile Leu Ala Ala Ser Phe Gly Leu
        835                 840                 845
Leu Ala Cys Ile Phe Phe Asn Lys Val Tyr Ile Ile Leu Phe Lys Pro
    850                 855                 860
Ser Arg Asn Thr Ile Glu Glu Val Arg Ser Ser Thr Ala Ala His Ala
865                 870                 875                 880
Phe Lys Val Ala Ala Arg Ala Thr Leu Arg Arg Pro Asn Ile Ser Arg
                885                 890                 895
Lys Arg Ser Ser Ser Leu Gly Gly Ser Thr Gly Ser Ile Pro Ser Ser
            900                 905                 910
Ser Ile Ser Ser Lys Ser Asn Ser Glu Asp Arg Phe Pro Gln Pro Glu
        915                 920                 925
Arg Gln Lys Gln Gln Gln Pro Leu Ser Leu Thr Gln Gln Glu Gln Gln
    930                 935                 940
Gln Gln Pro Leu Thr Leu His Pro Gln Gln Gln Gln Gln Pro Gln Gln
945                 950                 955                 960
Pro Arg Cys Lys Gln Lys Val Ile Phe Gly Ser Gly Thr Val Thr Phe
                965                 970                 975
Ser Leu Ser Phe Asp Glu Pro Gln Lys Asn Ala Met Ala His Arg Asn
            980                 985                 990
Ser Met Arg Gln Asn Ser Leu Glu Ala Gln Arg Ser Asn Asp Thr Leu
        995                1000                1005
Gly Arg His Gln Ala Leu Leu Pro Leu Gln Cys Ala Asp Ala Asp Ser
   1010                1015                1020
Glu Met Thr Ile Gln Glu Thr Gly Leu Gln Gly Pro Met Val Gly Asp
025                1030                1035                1040
His Gln Pro Glu Met Glu Ser Ser Asp Glu Met Ser Pro Ala Leu Val
                1045                1050                1055
Met Ser Thr Ser Arg Ser Phe Val Ile Ser Gly Gly Gly Ser Ser Val
```

1060 1065 1070
Thr Glu Asn Val Leu His Ser
1075

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: Modified Base
(B) LOCATION: 3...3
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 12...12
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 15...15
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 18...18
(D) OTHER INFORMATION: Inosine

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

BTNYAYCARR TNGCNMCNAA RGAYAC 26

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: Modified Base
(B) LOCATION: 6...6
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 9...9
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 12...12
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 18...18

(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 21...21
(D) OTHER INFORMATION: Inosine

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GYRTKNGCNR YNRCRTRNAC NRCRTT 26

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: Modified Base
(B) LOCATION: 3...3
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 9...9
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 12...12
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 13...13
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 24...24
(D) OTHER INFORMATION: Inosine

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

MRNTGYCCNK ANNAYMARTA YGCNAA 26

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: Modified Base
(B) LOCATION: 2...2
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 5...5
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 8...8
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 11...11
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 14...14
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 20...20
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 26...26
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 29...29
(D) OTHER INFORMATION: Inosine

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GNCKNAYNAR NATNAYRTAN MWYTTNGGNA C 31

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: Modified Base
(B) LOCATION: 3...3
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 6...6
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 9...9
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 12...12

(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 16...16
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 24...24
(D) OTHER INFORMATION: Inosine

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

ATNWSNYTNR TNTTYNGYTT YYTNTG 26

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: Modified Base
(B) LOCATION: 2...2
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 5...5
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 11...11
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 17...17
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 20...20
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 23...23
(D) OTHER INFORMATION: Inosine

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

RNATNSWRAA NAYYTCNACN RCNACCAT 28

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: Modified Base
(B) LOCATION: 6...6
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 9...9
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 12...12
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 15...15
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 21...21
(D) OTHER INFORMATION: Inosine

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GAYACNCCNA TNGTNAARGC NAAYAA 26

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: Modified Base
(B) LOCATION: 3...3
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 6...6
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 12...12
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 15...15
(D) OTHER INFORMATION: Inosine

(A) NAME/KEY: Modified Base
(B) LOCATION: 24...24
(D) OTHER INFORMATION: Inosine

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

AANGTNAYCC ANACNSWRCA RAANAC 26

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2550 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

ATGAAGCAGC TCTGCGCTTT CACTATTTCT TTGTTGTTTC TGAAGTTTTC TCTCATCCTG 60
TGCTGTTTGA CTGAACCAAG TTGCTTTTGG AGAATAAGGA ATAGTGAAGA TAGTGATGGA 120
GATTTACAAA GGGAATGTCA TTTTTACCTT TGGAAAACTG ATGAACCTAT TGAAGATAGT 180
TTTTATAATT ATGATTTAAG TTTTAGAATT GCAGCAAGTG AATATGAGTT TCTTCTCGTA 240
ATGTTTTTTG CTATCGATGA GATCAACAGG AATCCTTATC TTTTACCCAA CATAACTTTG 300
ATGTTCTCCT TCATTGGTGG AAACTGTCAG GATTTATTGA GAGTTATGGA CCAAGCATAT 360
ACACAAATAA ATGGACATAT GAATTTTGTT AATTATTTCT GTTATTTAGA TGATTCATGT 420
GCCATAGGTC TTACAGGACC ATCATGGAAA ACTTCCTTAA AACTGGCAAT GCACTCTTCG 480
ATGCCACTGG TTTTCTTTGG ACCATTTAAT CCTAACCTAC GCGACCATGA CCGGCTGCCC 540
CATGTCCATC AGGTAGCCCC CAAGGACACA CATTTGTCCC ATGGCATGGT CTCCTTGATG 600
TTTCACTTTA GATGGACTTG GATAGGACTG GTCATCTCAG ATGATGACCA GGGTATTCAG 660
TTTCTCTCAG ATTTAAGAGA AGAAAGCCAA AGGCATGGGA TCTGTTTAGC TTTTGTTAAT 720
ATGATCCCAG AAAACATGCA GATATACATG ACAAGGGCTA CAATATATGA TAAACACATT 780
ATGACATCTT CAGCAAAGGT TGTTATCATT TATGGTGAAA TGAACTCTAC TCTAGAAGCA 840
AGCTTTAGAA GATGGGAAGA GTTAGGTGCT CGGAGAATCT GGATCACAAC CTCACAATGG 900
GATGTCATCA CAAATAAAAA AGACTTCACC CTTAATCTCT TCCATGGGAT CATCACTTTT 960
GAACATCATA GATTTGAGAT TCCTAAATTA AATAAATTCA TGCAAACAAT GAACACTGCC 1020
AAATACCCAG TAGATATTTC TCATACTATA TTGGAGTGGA ATTATTTTAA TTGTTCAATA 1080
TCTAAGAACA GCATTAGAAT GCATCATATT ACATTCAACA ACACCTTGGA ATGGACATCA 1140
CTGCACAACT ATGATGTGGC GATGAGTGAT GAAGGTTACA ATTTGTACAA TGCTGTTTAT 1200
GCTGTGGCCC ACACCTACCA TGAATACATT TTTCAACAAG TAGAGTCTCA GAAAAAGGCA 1260
AAACCCAAAA GATATTTCAC TGCTTGTCAG CAGGTGTCTT CCTTGATGAA AACCAGGGTA 1320
TTTACGAACC CTGTTGGAGA ACTGGTGAAC ATGAAGCATA GGGAAAATCA GTGTACAGAG 1380
TATGATATTT TCATCATTTG GAATTTTCCA CAAGGCCTTG GATTAAAAGT GAAAATAGGA 1440
AGCTATTTAC CTTGTTTTCC ACAGAGACAA AAACTTCATA TATCTGATGA TTTGGAATGG 1500
GCCAAGGGAG GAACATCACC TCAGGTTCCC TCCTCCGTGT GTAGTGTGGC ATGTACTGCT 1560
GGATTCAGGA AAATTTATCA AAAAGAAACA GCAGACTGCT GCTTTGATTG TGTTCAGTGC 1620
CCAGAAAATG AGATTTCCAA CGAAACAGAT ATGGAACAGT GTGTGAGGTG TCCAGATGAT 1680
AAGTATGCCA ACATAGAGCA AACCCACTGC CTCTCAAGAG CTGTATCATT TCTGGCTTAT 1740
GAAGATTCAT TGGGGATGGC TCTAGGCTGC ATGGCACTGT CCTTCTCAGC CATCACAATT 1800
CTAATCCTCG TCACATTTGT GAAGTACAAA GATACTCCCA CTGTGAAGGC CAATAACCGC 1860
ATTCTCAGCT ACATCCTGCT CATCTCTCTC GTCTTCTGCT TTCTCTGCTC CCTGCTCTTC 1920
ATTGGACCTC CCGACCAGGT CACCTGCATC TTTCAGCAGA CCACATTTGG AGTATTGTTC 1980
ACTGTGTCTG TTTCTACAGT GTTGGCCAAA ACAATAACTG TGGTCATGGC TTTCAAGCTC 2040
ACTACTCCAG GAAGAAGGAT GAGAGGGATG ATGATGACAG GGGCACCTAA GTTGGTCATT 2100
CCCATTTGTA CCCTGATCCA ACTTGTTCTC TGTGGAATCT GGTTGGTCAC ATCTCCTCCC 2160
TTTATTGACA GAGACATACA ATCTGAGCAT GGGAAGATTG TCATTCTTTG CAATAAAGGC 2220
TCAGTCATTG CCTTCCACGT CGTCCTGGGA TACTTGGGCT CCTTGGCTCT GGGGAGCTTC 2280
ACGTTGGCTT TCCTGGCTAG GAACCTTCCT GACACATTCA ATGAAGCCAA GTTCCTAACT 2340
TTCAGCATGC TGGTGTTCTG CAGTGTCTGG ATCACCTTCC TCCCTGTCTA CCACAGCACC 2400
AGGGGGAGGG TCATGGTGGT TGTGGAGGTT TTCTCCATCT TGGCTTCTAG TGCAGGGTTG 2460
CTAATGTGTA TCTTTGTCCC AAAGTGTTAT GTTATTTTAA TTAGACCAGA TTCAAATTTT 2520
ATAAAGAACC ACAAAGGTAA ATTGCTTTAT 2550

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2424 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

ATGAAGCAGC TCTGCACTTT CACTATTTCA TTGTTGTTTC TGAAGTTTTC TCTCATCTTG 60
TGCTGTTGGA GTGAACCAAG CTGCTTTTGG AGGATAAAGA AGAGTGAAGA TAATGATGGA 120
GATTTACAAA GGGAGTGTCA TTTTTACCTT TGGAAAACTG ATGAACCTAT TGAAGATAGT 180
TTTTATAATT ATGATTTAAG TTTTAGAATT GCAGGAAGTG AATATGAGCT TCTTCTGGTA 240
ATGTTTTTTG CTACTGATGA GATCAACAAG AATCCTTATC TTTTACCCAA CATGAGTTTG 300
ATGTTCTCCA TCATTGGTGG AAACTGTCAT GATTTATTGA GAAGTCTGGA TCAAGAATAT 360
GCACAAATAG ATGGACATAT GAATTTTGTT AATTATTTCT GTTATTTAGA TGATTCATGT 420
GCCACAGGCC TTACAGGACC ATCATGGAAA ACATCCTTAA AACTGGCAAT GCATTCTTCA 480
ATGCCACTGG TTTTCTTTGG ACCATTTAAT CCTAACCTAC GCGACCATGA CCGGCTGCCC 540
CATGTCCATC AGGTAGCCCC CAAGGACACA CATTTGTCCC ATGGCATGGT CTCCTTGATG 600
TTTCATTTTA GGTGGACTTG GATAGGACTG GTCATCTCAG ATGATGATCA GGGTATTCAG 660
TTTCTCTCAG ATTTAAGAGA AGAAAGCCAA AGGCATGGGA TCTGTTTGGC TTTTGTTAAT 720
ATGATCCCAG AAAACATGCA GATATACATG ACAAGGGCTA CAATATATGA TACACAAATT 780
ATGACATCTT CAGCAAAGGT TGTTATCATT TATGGTGACA TGAACTCTAC TCTAGAAGCA 840
AGCTTTAGAA GATGGGAAGA GTTAGGTGCT CGGAGAATCT GGATCACAAC CACACAATGG 900
GATGTCATCA CAAATAAAAA AGACTTCACC CTTAATCTCT TCCATGGGAC TATTACTTTT 960
GCACACCACA AAGATGAGAT TCCTAAATTT AGGAATTTTA TGCAAACAAA GAAAACTGCC 1020
AAATACCTTG TAGATATTTC TCATACTATT TTGGAGTGGA ATTATTTTAA TTGTTCAATC 1080
TCTAAGAACA GCAGTAAAAT GGGTCATTTT ACATTCAACA ACACATTGCA ATGGACAGCA 1140
CTGCACAACT ATGATATGGC CCTGAGCGAT GAAGGTTACA ATTTGTATAA TGCTGTTTAT 1200
GCTGTGGCCC ACACCTACCA TGAATACATT CTTCAACAAG TAGAGTCTCA GAAAAAGGCA 1260
AAACCCAAAA GATATTTCAC TGCTTGTCAG CAGGTGTCTT CCTTGATGAA AACCAGGGTA 1320
TTTATGAACC CTGTTGGAGA ACTGGTGAAC ATGAAGCATA GGGAAAATCA GTGTACAGAG 1380
TATGATATTT TCATCATTTG GAATTTTCCA CAAGGCCTTG GATTAAAAGT GAAAGTAGGA 1440
AGCTATTTAC CTTGCTTTCC AAAGAGTCAA CAACTTCATA TAGCTGATGA TTTGGAATGG 1500
GCCATGGGAG GAACATCAGT GGATATGGAA CAGTGTGTGA GATGTCCAGA TAATAAATAT 1560
GCCAATTTAG AGCAAACCCA CTGCCTCCAA AGAACGGTGT CATTTCTGGC TTATGAAGAT 1620
CCATTGGGGA TGGCTCTAGG CTGCATGGCA CTGTCCTTCT CGGCCATCAC AATTCTAGTC 1680
CTCGTCACAT TTGTGAAGTA CAAGGATACT CCCATTGTGA AGGCCAATAA CCGCATTCTC 1740
AGCTACATCC TGCTCATCTC TCTCGTCTTC TGCTTTCTCT GTTCCCTGCT CTTCATTGGA 1800
CATCCCGACC AGGTCACCTG CATCTTGCAG CAGACCACAT TTGGAGTATT GTTCACTGTG 1860
TCTGTTTCTA CAGTGTTGGC CAAAACAATA ACTGTGGTCA TGGCTTTCAA GCTCACTACT 1920
CCAGGAAGAA GGATGAGAGG GATGATGATG ACAGGGGCAC CTAAGTTGGT CATTCCCATT 1980
TGTACCCTGA TCCAACTTGT TCTCTGTGGA ATCTGGTTGG TCACATCTCC TCCCTTTATT 2040
GACAGAGATA TACAATCTGA ACATGGGAAG ATTGTCATTC TTTGCAATAA AGGCTCTGTC 2100
GTTGCCTTCC ACGTCGTCCT GGGATACTTG GGCTCCTTGG CTCTGGGGAG CTTCACTTTG 2160
GCTTTCTTGG CTAGGAACCT TCCTGACACA TTCAATGAAG CCAAGTTCCT AACTTTCAGC 2220
ATGCTGGTGT TCTGCAGTGT CTGGATCACC TTCCTCCCTG TCTACCACAG CACCAGGGGG 2280
AAGGTCATGG TGGTTGTGGA GGTTTTCTCC ATCTTGGCTT CTAGTGCAGG GTTGCTAATG 2340
TGTATCTTTG TCCCAAAGTG TTATGTTATT TTAATTAGAC CAGATTCAAA TTTTATACAG 2400
AACCACAAAG GTAAATTGCT TTAT 2424

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2409 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

```
CATTTTTACC TTGGGGCAGT TGATAAACCA ATTGAAGATA ATTTTTATAA TTCACTTTTA     60
AAGTTTAGAA TTGCAGCAAG TGAATATGAG TTTCTTCTGG TAATGTTTTT TGCTACTGAT    120
GAGATCAACA AGAATCCTTA TCTTTTACCC AACATAACTT TGATGTTCTC CATCATTGGT    180
GGAAACTGTC ATGATTTATT GAGAGGTTTG GATCAAGCAT ATACACAAAT AAATGGACAT    240
ATGAATTTTG TTAATTATTT CTGTTATTTA GATGATTCAT GTGCCATAGG TCTTACAGGA    300
CCATCATGGA AAACATCCTT AAATCTGGCA ATGCATTCTT CAATGCCACT GGTTTTCTTT    360
GGATCATTTA ATCCTAACCT ACATGACCAT GACCGGCTGC ACCATGTCCA TCAAGTAGCC    420
ACCAAGGACA CACATTTGTC CCATGGCATT GTCTCCTTGA TGTTTCATTT TAGATGGACT    480
TGGATAGGAC TGGTCATCTC AGATGATGAC AAGGGTATTC AGTTTCTCTC AGATTTAAGA    540
GAAGAAAGCC AAAGGCATGG GATCTGTTTA GCTTTTGTTA ATATGATCCC AGAAAACATG    600
CAGATATACA TGACAAGGGC TACAATATAT GATAAACAAA TTATGACGTC TTTAGCAAAA    660
GTTGTTATCA TTTATGGTGA AATGAACTCT ACACTAGAAG TAAGCTTTAG AAGATGGGAA    720
AATTTAGGTG CTCGGAGAAT CTGGATCACA ACCTCACAAT GGGATGTCAT CACAAATAAA    780
AAAGAATTCA CCCTTAATCT CTTCCATGGG ACTATTACTT TTGCACACCG CAGATTTGAG    840
ATTCCTAAAT TTAAAAAATT TATGCAAACA ATGAACACTG CCAAATACCC AGTAGATATT    900
TCTCATACTA TATTGGAGTG GAATTATTTT AATTGTTCAA TCTCTAAGAA CAGCAGTAAA    960
ATGGATCATA TTACATTCAA CAACACATTG GAATGGACAG CACTGCACAA CTATGATATG   1020
GTGATGAGTG ATGAAGGTTA CAATTTGTAT AATGCTGTTT ATGCTGTGGC CCACACCTAC   1080
CATGAACATA TTTTTCAACA AGTAGAGTCT CAGAAAAAGG CAAAACCCAA AAGATTTTTC   1140
ACTGTTTGTC AGCAGGTGTC TTCCTTGATG AAAACCAGGG TATTTACTAA CCCTGTTGGA   1200
GAACTGGTGA ACATGAAGCA TAGGGAAAAT CAGTGTACAG AGTATGACAT TTTCCTCATT   1260
TGGAACTTTC CACAAGGCCT TGGATTAAAA GTGAAAATAG GAAGCTATTT ACCTTGTTTT   1320
CCACAGAGAC AAGAACTTCA TATATCTGAT GATTTGGAAT GGGCCATGGG AGGAACATCA   1380
GTGGTTCCCT CCTCTGTGTG TAGTGTGGCA TGTACTGCAG GATTCAGGAA AATTCATCAG   1440
AAAGAAACAG CAGACTGCTG CTTTGATTGT GTTCAGTGCC CAGAAAATGA GGTTTCCAAT   1500
GAAACAGATA TGGAACAGTG TGTGAAGTGT CCATATGATA AGTATGCCAA CATAGAGAAA   1560
ACCCACTGCC TCTCAAGAGC TGTATCATTT CTGGCTTATG AAGATCCATT GGGGATAGCT   1620
CTAGGCTGCA TAGCACTGTC CTTCTCAGCC ATCACAATTC TAGTACTAAT CACATTTTTG   1680
AAGTACAAGG ATACTCCCAT TGTGAAGGCC AATAACCGCA TTCTCAGCTA CATCCTGCTC   1740
ATCTCTCTAG TCTTCTGCTT TCTCTGCTCC CTGCTCTTCA TTGGACATCC AAACCAGGTC   1800
TCCTGCGTCT TGCAGCAGAC CACATTTGGA GTATTTTTCA CTGTGTCTGT TTCTACAGTG   1860
TTGGCCAAAA CAATAACTGT GGTCATGGCT TTCAAGCTCA CTACTCCAGG AAGAAGAATG   1920
AGAGAGATGT TGGTAACAGG GGCACCTAAG TTGGTCATTC CCATTTGTAC CCTAATCCAA   1980
TTTGTTCTCT GTGGAATCTG GTTGATAACA TCTCCTCCAT TTATTGACAG AGATATACAA   2040
TCTGAGCATG GGAAGATTGT CATTCTTTGC AATAAAGGCT CTGTCATTGC CTTCCATGTT   2100
GTCCTGGGAT ACTTGGGCTC CTTGGCTCTG GGGAGCTTCA CTTTGGCTTT CTTGGCTAGG   2160
AACCTTCCTG ACACATTCAA TGAAGCCAAA TTCCTGACTT TCAGCATGCT GGTGTTCTGC   2220
AGTGTCTGGA TCACCTTTCT CCCTGTCTAC CATAGCACCA GGGGGAAGGT CATGGTGGTT   2280
GTGGAGGTTT TCTCAATCTT GGCTTCTAGT GCAGGGTTGC TAATGTGTAT CTTTGTCCCA   2340
AAGTGTTATG TTATTTTAGT TAGACCAGAT TCAAATTTTA TACGGAAGTA CAAAGATAAA   2400
TTTCGTTAT                                                           2409
```

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2556 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

```
ATGTTCATTT TCATGGGAGT CTTCTTCCTA CTTAATATTA CACTTCTCAT GGCCAATTTC     60
ATTGATCCCA GGTGCTTTTG GAGAATAAAT TTGGATGAAA TAACGGATGA ATATTTGGGA    120
TTATCTTGTG CTTTCATCCT GGCAGCTGTT CAGACACCCA TTGAAAAAGA TTATTTCAAC    180
ACGACTCTTA ATTTTCTAAA AACTACTAAA AACCACAAAT ATGCTTTGGC ATTGGTGTTT    240
GCAATGGATG AAATCAACAG ATATCCTGAT CTTTTACCAA ATATGTCTTT GATTATCAGA    300
TACTCTTTGG GCCATTGTGA TGGAAAAACT GTAACACCTA CACCATATTT ATTTCATAGA    360
AAAAAGCAAA GCCCTATTCC TAATTATTTC TGTAATGAAG AGAGTATGTG TTCATTTCTG    420
CTTTCAGGAC CCAATTGGGA TGAATCTTTA AGTTTCTGGA AGTACCTGGA CAGCTTCTTA    480
TCTCCACGTA TCCTTCAGCT TTCCTATGGA TCTTTCAGTT CCATCTTCAG TGATGATGAA    540
CAATATCCCT ATCTCTATCA GATGGCCCCA AAAGACACAT CTCTAGCATT GGCAATGGTC    600
TCCTTCATAC TTTATTTGAA ATGGAATTGG ATTGGCCTTG TCATCCCAGA TGATGATCAA    660
```

GGAAACCAAT TTCTTTTAGA GTTGAAGAAA CAGAGTGAAA ACAAAGAAAT TTGCTTTGCC 720
TTTGTGAAAA TGATCTCTGT TGATGAAGTT TCATTTCCAC AAAAAACTGA AATAAACTAC 780
AAACAAATTG TGAAGTCACT AACAAATGTT ATTATCATTT ATGGAGAAAC ATATAATTTC 840
ATTGATTTGA TCTTCAGAAT GTGGGAACCT CCCATTTTAC AGAGAATATG GATCACCACA 900
AAACAATTGA ATTTCCCTAC CAGTAAGACA GACATAAGTC ATGACACATT CTATGGATCA 960
CTTACTTTTC TACCCCACCA TGGTGAGATT TCTGGCTTTA AAAATTTTGT ACAGACATGG 1020
TTCCATCTCA GAAACACAGA TTTATGTCTA GTAATGCCAG AGTGGAAATA TATTAACTCT 1080
GAAGACTCAG CATCTAATTG TAAAATACTT AAGAACAGTT CATCTGATGC CTCATTTGAT 1140
TGGCTAATGG AAGAGAAGCT TGACATGGCC TTTAGTGAGA ATAGTCATAA CATATATAAT 1200
GCTGTGCATG CCATAGCCCA TGCCCTCCAT GAGATGAATC TGCAACAGGC TGATAATCAG 1260
GCAATAGATA ATGGAAAAGG AGCCAGTTCT CACTGCTTGA AGGTAAACTC CTTTCTAAGA 1320
AGGACCTACT TCACTAATCC TCTTGGGGAC AAAGTGTTTA TGAAGCAAAG AGTAATAATG 1380
CAGGATGAAT ATGACATTGT TCACTTTGCG AATCTCTCAC AACACCTTGG GATTAAGATG 1440
AAGTTAGGAA AGTTCAGCCC ATATTTACCA CATGGTCGAC ACTCTCACTT ATACGTAGAC 1500
ATGATTGAGT TGGCCACAGG AAGAAGAAAG ATGCCATCCT CTGTGTGCAG TGCAGATTGT 1560
AGTCCTGGAT TCAGAAGATT ATGGAAGGAG GGAATGGCAG CCTGCTGTTT TGTTTGCAGC 1620
CCCTGCCCTG AAAATGAAAT TTCTAATGAG ACAAATATGG ATCAATGCGT GAATTGTCCA 1680
GAATACCAAT ATGCCAACAC AGAACAGAAC AAATGTATTC AGAAAGGTGT CACCTTCCTA 1740
AGCTATGAAG ACCCCTTGGG GATGGCACTT GCCTTAATGG CCTTCTGCTT CTCTGCATTC 1800
ACAGCTGTGG TACTTTGTGT CTTTGTGAAG CACCATGACA CTCCTATTGT GAAGGCCAAT 1860
AACAGAAGCC TCAGCTATCT ATTACTCATG TCACTCATGT TCTGTTTTCT GTGCTCCTTT 1920
TTCTTCATTG GCCTTCCAAA CAAAGTCATC TGTGTCTTAC AGCAAATCAC ATTTGGAATT 1980
GTATTCACTG TGGCTGTTTC CACAGTTCTG GCCAAAACAG TCACTGTGGT TCTAGCTTTC 2040
AAAGTCACAG TCCCAGGAAG AAGATTGAGA TACTTCCTTG TATCAGGGAC ACTAAACTAC 2100
ATTATTCCTA TATGTTCCCT ACTCCAATGT GTTCTGTGTG CAATCTGGCT AGCAGTCTCT 2160
CCTCCCTTTG TTGATATTGA TGAACACTCT CAGCATGGCC ACATCATCAT TGTGTGCAAC 2220
AAGGGCTCAG TTACTGCATT CTACTGTGTC CTTGGATACT TGGCCTGCCT GGCACTGGGA 2280
AGCTTCACTT TGGCTTTCTT GGCCAAGAAT CTGCCTGATG CATTCAATGA AGCCAAGTTC 2340
TTGACCTTCA GCATGCTAGT GTTCTGCAGT GTCTGGGTCA CCTTCCTCCC TGTGTACCAT 2400
AGCACAAAGG GCAAACACAT GGTTGCTGTG GAGATCTTCT CTATCTTGGC ATCCAGTGCA 2460
GGGATGCTTG GATGTATTTT TGTACCCAAG ATTTATATCA TTTTAATGAG ACCAGAGAGA 2520
AATTCTACCC AAAAGATCAG AGAAAAATCA TATTTT 2556

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2169 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

ATCTGTAATG AAGAGAGTAT GTGTTCATTT CTGCTTTCAG GACCCAATTG GGATGAATCT 60
TTAAGTTTCT GGAAGTACCT GGACAGCTTC TTATCTCCAC ATATCCTTCA GCTTTCCTAT 120
GGATCTTTCA GTTCCATCTT CAGTGATGAT GAACAATATC CCTATCTCTA TCAGATGGCC 180
CCAAAGGACA CATCTCTAGC ATTGGCAATG GTCTCCTTCA TACTTTATTT GAAATGGAAT 240
TGGATTGGCC TTGTCATCCC AGATGACGAT CAAGGAAACC AATTTCTTTT AGAGTTGAAG 300
AAACAGAGTG AAAACAAAGA AATTTGCTTT GCCTTTGTGA AAATGATATC TGTTGATGAA 360
GTTTCATTTC CACAAAAAAC TGAAATATAC TACAAACAAA TTGTGAAGTC ATTAACAAAT 420
GTTATTATCA TTTATGGAGA AACATATAAT TTCATTGATT TGATCTTCAG AATGTGGGAA 480
CCTCCCATTT TACAGAGAAT ATGGATCACC ACAAAACAAT TGAATTTCCC TACCAGTAAG 540
ACAGACATAA GTCATGACAC ATTCTATGGA TCACTTACTT TTCTACCCCA CCATGGTGAG 600
ATTTCTGGCT TTAAAAATTT TGTACAGACA TGGTTCCATC TCAGAAACAC AGATTTATAT 660
CTAGTAATGC CAGAGTGGAA ATATATTAAC TCTGAAGACT CAGCATCTAA TTGTAAAATA 720
CTGAAGAACA GTTCATCTGA TGCCTCATTT GATTGGCTAA TGGAACAGAA GCTTGACATG 780
GCCTTTAGTG ATAATAGTCA TAACATATAT AATGTTGTGC ATGCCATAGC CCATGCCCTC 840
CATGAGATGA ATCTGCAACA GGCTGATAAT CAGGCAATAG ATAATGGAAA AGGAGCCAGT 900
TCTCACTGCT TGAAGGTAAA CTCCTTTCTA AGAAGGACCT ACTTCACTAA TCCTCTTGGG 960
GACAAAGTGT TTATGAAGCA AAGAGTAATA ATGCAGGATG AATATGACAT TGTTCACTTT 1020
GCGAATCTCT CACAACACCT TGGGATTAAG ATGAAGTTAG GAAAGTTCAG CCCATATTTA 1080
CCACATGGTC GACACTCTCA CTTATACGTA GACATGATTG AGTTGGCCAC AGGAAGAAGA 1140
AAGATGCCAT CCTCTGTGTG CAGTGCAGAT TGTAGTCCTG GATTCAGAAG ATTATGGAAG 1200

GAGGGAATGG CAGCCTGCTG TTTTGTTTGC AGCCCCTGCC CTGAAAATGA AATTTCTAAT 1260
GAGACAAATA TGGATCAATG CGTGAATTGT CCAGAATACC AATATGCCAA CACAGAACAG 1320
AACAAATGTA TTCAGAAAGG TGTCACCTTC CTAAGCTATG AAGACCCCTT GGGGATGGCA 1380
CTTGCCTTAA TGGCCTTCTG CTTCTCTGCA TTCACAGCTG TGGTACTTTG TGTCTTTGTG 1440
AAGCACCATG ACACTCCTAT TGTGAAGGCC AATAACAGAA GCCTCAGCTA TCTATTACTC 1500
ATGTCACTCA TGTTCTGTTT TCTGTGCTCC TTTTTCTTCA TTGGCCTTCC AAACAAAGTC 1560
ATCTGTGTCT TACAGCAGAT CACATTTGGA ATTGTATTTA CTGTAGCTGT TTCCACAGTT 1620
CTGGCCAAAA CAGTCACTGT GGTTCTAGCT TTCAAAGTCA CAGACCCAGG AAGAAGATTG 1680
AGATACTTCC TTGTATCAGG GACACTAAAC TACATTATTC CTATATGTTC CCTACTCCAA 1740
TGTGTTCTGT GTGCAATCTG GCTAGCAGTC TCTCCTCCCT TTGTTGATAT TGATGAACAC 1800
TCTCAGCATG GCCACATCAT CATTGTGTGC AACAAGGGCT CAGTTACTGC ATTCTACTGT 1860
GTCCTTGGAT ACTTGGCCTG CCTGGCACTG GGAAGCTTCA CTTTGGCTTT CTTGGCCAAG 1920
AATCTGCCTG ATGCATTCAA TGAAGCCAAG TTCTTGACCT TCAGCATGCT AGTGTTCTGC 1980
AGTGTCTGGG TCACCTTCCT CCCTGTGTAC CATAGCACAA AGGGCAAACA CATGGTTGCT 2040
GTGGAGATCT TCTCCATCTT GGCATCCAGT GCAGGGATGC TTGAATGTAT TTTTGTACCC 2100
AAGATTTATA TCATTTTAAT GAGACCAGAG AGAAATTCTA CCCAAAAGAT CAGGGAAAAA 2160
TCATATTTC 2169

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1889 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

GAATTCGGCT TCTGCACCAA ATGGCGACGA AAGACACATC TCTTTCACTT GCCATTGTTT 60
CTTTGATGGT TCATTTTAGG TGGTCTTGGG TTGGTCTAAT TCTCCCAGAT GACCACAAAG 120
GAAATAAAAT ACTATCAGAT TTTAGAAAGG AGATGGAAAG AAAAAGAATC TGTACGGCTT 180
TTGTAAAAAT GATTCCTGCC ACATGGACTT CATCTTTTGT CAAATTCTGG GAAAATATGG 240
ATGACACCAA CATAATAATT ATTTATGGTG ACATTGATTC TCTAGAAGGT CTAATGCGAA 300
ATATTGGGCA AAGGTTATTG ACATGGCATG TCTGGGTCAT GAACATTGAA CCCCATATTA 360
TTGAATATGA TAATTATTTC ATGTTAGATT CATTCCATGG AAGTTTAATT TTTAAGCACA 420
ATTATAGAGA GAATTTTGAG TTTACCAAAT TTATTCGAAC AGTTAATCCT AAAAAATACC 480
CAGAAGACAT TTATCTCCCT AAGATGTGGT ATTTGTTCTT CATGTGCTCA TTTTCTGATA 540
TTAATTGTCA AGTTTTGGAC AGCTGTCAAA CAAATGCTTC TTTGGATATG TTACCTAGTC 600
AGATATTTGA TGTGGTCATG AGTGAAGAGA GCACAAGTAT TTACAATGCT GTGTACGCTG 660
TGGCTCACAG CCTCCATGAG ATGAGACTTC AGCAACTTCA AACACAACCG TGTGAAAATG 720
AAGAAGGGAT GGAGTTCTTT CCATGGCAGC TTAATACTTT CCTGAAGGAT ATTGAGGTGA 780
GAGTCAACAG TTTAGACTGG AGACAGAGAA TAGATGCTGA ATATGACATT CTTAACCTCT 840
GGAATTTACC AAAGGGTCTT GGACTAAAAG TGAAAATAGG AAACTTTTAT GCAAATGCTC 900
CCCAGGGTCA ACAATTGTCT TTATCTGAAC AGATGATTCA ATGGCCAGAA ATATTTTCAG 960
AGATCCCTCA GTCGGTGTGC AGTGAGAGTT GTGGGCCTGG ATTCAGGAAA GTAACCCTGG 1020
AGAATAAGGC TATCTGCTGC TACAATTGTA CTCCCTGTGC AGACAATGAG ATTTCTAATG 1080
AGACAGATGT AGACCAGTGT GTGAAGTGTC CAGAGAGTCA TTATGCAAAT ACAGAGAAGA 1140
GCAACTGCTA TCAAAAGTCT GTGAGCTTTC TGGGCTATGA AGACCCTTTG GGGATGGCTC 1200
TAGCCAGCAT AGCTTTGTGC TTGTCTGCAC TAACTGCCTT TGTTATTGGC ATATTTGTGA 1260
AACACAAAGA CACTCCTATT GTTAAGGCCA ATAATCAAGC TCTGAGTTAC ACTTTGCTCA 1320
TCACACTCAA ATTCTGTTTC CTATGTTCTT TGAACTTCAT TGGTCAGCCC AACACAGTTG 1380
CCTGCATCCT TCAGCAGACC ACCTTTGCAG TTGCTTTCAC TATGGCTCTT GCCACTGTGT 1440
TGGCCAAAGC TATCACTGTG GTTCTTGCCT TTAAGGTCAG TTTTCCAGGG AGAATGGTAA 1500
GATGGCTAAT GATATCAAGG GGTCCAAACT ATATCATTCC TATCTGCACC CTGATCCAAC 1560
TTCTTCTTTG TGGAATATGG ATGGCAATAT CTCCACCATA CATTGACCAA GATGCTCATA 1620
TTGAACATGG TCACATCATC ATTTTGTGCA ACAAGGGCTC AGCTGTTGCC TTCCACTCTG 1680
TCCTGGGATA CCTCTGCTTC TTGGCCCTTG GGAGTTATAC CATGGCCTTC TTGTCAAGAA 1740
ATTTGCCTGA TACATTCAAC GAATCCAAAT TTATCTCACT AAGTATGCTG GTATTCTTCT 1800
GTGTCTGGAT CACCTTTCTT CCTGTCTACC ACAGCACTAA AGGGAAGGTC ATGGTCGCCG 1860
TCGAGGTCTT TTGCATCCAA GCCGAATTC 1889

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1889 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

GAATTCGGCT TCTGCATCAA ATGGCGACGA AGGACACATC TCTTTCACTT GCCATTGTTT 60
CTTTGATGGT TCATTTTAGG TGGTCTTGGG TTGGTCTAAT TCTCCCAGAT GACCACAAAG 120
GAAATAAAAT ACTATCAGAT TTTAGAAAGG AGATGGAGAG AAAAAGAATC TGTACGGCTT 180
TTGTAAAAAT GATTCCTGCC ACATGGACTT CATCTTTTGT CAAATTCTGG GAAAATATGG 240
ATGACACCAA CATAATAATT ATTTATGGTG ACATTGATTC TCTAGAAGGT CCAATGCGAA 300
ATATTGGGCA AAGGTTATTG ACATGGCATG TCTGGGTCAT GAACATTGAA CCCCATATTA 360
TTGAATATGA TAATTATTTC ATGTTAGATT CATTCCATGG AAGTTTAATT TTTAAGCACA 420
ATTATAGAGA GAATTTTGAG TTTACCAAAT TTATTCGAAC AGTTAATCCT AAAAAATACC 480
CAGAAGACAT TTATCTCCCT AAGATGTGGT ATTTGTTCTT CATGTGCTCA TTTTCTGATA 540
TTAATTGTCA AGTTTTGGAC AGCTGTCAAA CAAATGCTTC TTTGGATATG TTACCTAGTC 600
AGATATTTGA TGTGGTCATG AGTGAAGAGA GCACAAGTAT TTACAATGCT GTGTACGCTG 660
TGGCTCACAG CCTCCATGAG ATGAGACTTC AGCAACTTCA AACACAACCG TGTGAAAATG 720
AAGAAGGGAT GGAGTTCTTT CCATGGCAGC TTAATACTTT CCTGAAGGAT ATTGAGGTGA 780
GAGTCAACAG TTTGGACTGG AGACAGAGAA TAGATGCTGA ATATGACATT CTTAACCTCT 840
GGAATTTACC AAAGGGTCTT GGACTAAAAG TGAAAATAGG AAACTTTTAT GCAAATGCTC 900
CCCAGGGTCA ACAATTGTCT TTATCTGAAC AGATGATTCA ATGGCCAGAA ATATTTTCAG 960
AAGTCCCTCA GTCTGTGTGC AGTGAGAGTT GTAGGCCTGG ATTCAGGAAA GTATCCCTGG 1020
ATGATAAGGC CATCTGCTGC TACAAGTGCA CTCCTTGTGC CGACAATGAG ATATCTAATG 1080
AGACAGATGT AGACCAGTGT GTGAAGTGTC CAGAGAGTCA TTATGCAAAT ACAGAGAAGA 1140
GCAACTGCTT CCCAAAATCT GTGAGCTTTC TGGCCTATGA AGACCCCTTG GGGATGGCTC 1200
TAGCCAGCAT AGCTTTGTGC TTATCTGCAC TCACTGTCTT TGTTATTGGC ATCTTTGTGA 1260
AAAACAGAGA CACTCCTATT GTCAAGGCCA ATAATCGGAC TCTAAGTTAC ATTTTGCTCA 1320
TCACACTCAC CTTTTGTTTC TTATGTTCTT TGAACTTCAT TGGTCAGCCC AACACAGCTG 1380
CCTGCATCCT TCAGCAGACC ACCTTTGCAG TTGCTTTCAC TATGGCTCTT GCCACTGTGT 1440
TGGCCAAAGC TATTACTGTA GTCCTTGCCT TTAAGATCAG TTTTCCAGGG AGAATGTTAA 1500
GGTGGCTAAT GATATCAAGG GGTCCAAGAT ACATCATTCC TATCTGCACA CTGATCCAGC 1560
TTCTTCTTTG TGGAATATGG ATGGCAACTT CTCCACCATT CATTGACCAA GATGTTAATA 1620
CTGAAGATGG ATACATCATC CTTTTGTGCA ACAAGGGCTC AGCTGTTGCC TTCCATTCAG 1680
TCCTGGGATA CCTCTGTTTC TTGGCCCTTG GGAGTTATAC CATGGCCTTC TTGTCTAGAA 1740
ATTTGCCTGA TACATTCAAT GAATCCAAAT TTCTGTCATT CAGTATGCTG GTGTTCTTCT 1800
GTGTCTGGGT CACCTTTCTT CCTGTCTACC ACAGCACTAA AGGGAAAGTT ATGGTCGTCG 1860
TCGAAGTCTT CTGCATCCAA GCCGAATTC 1889

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 270 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

ATGAAGAAGC TCTGTGCTTT CACGATTTCA TTGTTGTTTC TGAAGTTTTC TCTCATCTTG 60
TGCTGTTGGA GTGAACCAAG TTGCTTTTGG AGGATAAAGA ATAGTGATGA TAATGACGGA 120
GATTTGCAAA GGGAATGTCA TTTTTACCTT GGGGCAGCTG ATACACCAGT TGAAGATAAT 180
TTTTATAGTT CACTTTTAAA ATTTAGGTTT TCTTTGGACC ATTTAATCCT AACCTACGCG 240
ACCATGACCG GCTGCCCCAT GTCCATCAGG 270

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1308 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

```
ATGAAGAAGC TCTGTGCTTT CACGATTTCA TTGTTGTTTC TGAAGTTTTC TCTCATCTTG     60
TGCTGTTGGA GTGAACCAAG TTGCTTTTGG AGGATAAAGA ATAGTGATGA TAATGACGGA    120
GATTTGCAAA GGGAATGTCA TTTTTACCTT GGGGCAGCTG ATACACCAGT TGAAGATAAT    180
TTTTATAGTT CACTTTTAAA ATTTAGAATT GCAGCAAGTG AATATGAGTT TCTTCTCGTA    240
ATGTTTTTTG CTATCGATGA GATCAACAGG AATCCTTATC TTTTACCCAA CATAACTTTG    300
ATGTTCTCCT TCATTGGTGG AAACTGTCAG GATTTATTGA GAGTTATGGA CCAAGCATAT    360
ACACAAATAA ATGGACATAT GAATTTTGTT AATTATTTCT GTTATTTAGA TGATTCATGT    420
GCCATAGGTC TTACAGGACC ATCATGGAAA ACTTCCTTAA AACTGGCAAT GCACTCTTCG    480
ATGCCACTGG TTTTCTTTGG ACCATTTAAT CCTAACCTAC GCGACCATGA CCGGCTGCCC    540
CATGTCCATC AGGTAGCCCC CAAGGACACA CATTTGTCCC ATGGCATGGT CTCCTTGATG    600
TTTCACTTTA GATGGACTTG GATAGGAATG GTCATCTCAG ATGATGACCA GGGTATTCAG    660
TTTCTCTCAG ATTTAAGAGA AGAAAGCCAA AGGCATGGGA TCTGTTTAGC TTTTGTTAAT    720
ATGATCCCAG AAAACATGCA GATATACATG ACAAGGGCTA CAATATATGA TCAACAAATT    780
ATGACATCTT CAGCAAAGGT TGTTATCATT TATGGTGAAA TGAACTCTAC TCTAGAAGTA    840
AGCTTTAGAA GATGGGAAGA GTTAGGTGCT CGGAGAATCT GGATCACAAC CTCACAATGG    900
GATGTCATCA CAAATAAAAA AGACTTCACC CTTAATCTCT TCCATGGGAC TATCACTTTT    960
GCACACCACA GAGTTGAGAT TCCTAAATTA AATAAATTCA TGCAAACAAT GAACACTGCC   1020
AAATACCCAG TAGATATTTC TCATACTATA TTGGAGTGGA ATTATTTTAA TTGTTCAATA   1080
TCTAAGAACA GCATTAGAAT GCATCATATT ACATTCAACA ACACCTTGGA ATGGACATCA   1140
CTGCACAACT ATGATATGGC GATGAGTGAT GAAGGTTACA GTTTATATAA TGCTGTTTAT   1200
GCTGTGGCCC ACACCTACCA TGAATACATT TTTCAACAAG TAGAGTCTCA GAAAAAGGCA   1260
AAACCCAAAA GATATTTCAC TGCTTGTCAG CAGATATGGA ACAGTGTG                1308
```

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1296 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

```
ATGAAGAAGC TCTGTGCTTT CACTATTTCA TTTTTGTCTC TGAAGTTTTC TCTCATCTTG     60
TGCTGTTTGA CTGAAGCAAG TTGCTTTTGG AGGATAAAGA ATAGTGAAGA TAGTGATGGA    120
GATTTGCAAA GAGAATGTCA TTTTTACCTT TGGGTAATTG ATAAACCTAT TGAAGATAAT    180
TTTTATAATT CAGTTTTAAA TTTTAGAATA TCAGCAAGTG AATATGAGTT TCTTCTGGTA    240
ATGTTTTTTG CTACTGATGA GATCAACAAG AATCCTTATC TTTTACCCAA CATAACTTTG    300
ATATTCAGCA TCGTTGGTGG TCACTGTCAT GATTTATTGA GAGGTCTGGA TCAATCATAT    360
ACACAAATAA ATGGACGTGT GAATTTTGTT AATTATTTCT GTTATTTAGA TGATTCATGT    420
AACATAGGCC TTACAGGACC ATCATGGAAA AAATCCTTAA AACTGGCAAT GGATTCTTCA    480
ATACCAATGG TTTTCTTTGG ACCATTTAAT CCTAACCTAC GCGACCATGA CCGGCTGCCC    540
CATGTCCATC AGGTAGCCCC CAAGGACACA CATTTATCCC ATGGCATGGT CTCCTTGATG    600
TTTCATTTTA GATGGACTTG GATAGGACTG GTCATCTCAG ATGATGACCA GGGTATTCAG    660
TTTCTCTCAG ATTTAAGAGA AGAAAGCCAA AGGCATGGGA TCTGTTTAGC TTTTGTTAAT    720
ATGATCCCAG AAAACATGCA GATATACATG ACAAGGGCTA CAATATATGA TAAACAAATT    780
ATGACATCTT CAGCAAAGGT TGTTATCATT TATGGTGAAA TGAACTCTAC TCTAGAAGTA    840
AGCTTCAGAA GATGGGAAGA TTTAGGTGCT CGGAGAATCT GGATCACAAC CTCACAATGG    900
GATATCATAT TAAATAAAAA AGAATTCACT CTTAATCTCT TCCATGGCCC TATCACTTTT    960
GCACACCACA AAGTTGAGAT TCCTAAATTA AGGAATTTTA TGCAAACAAT GAACACTGCC   1020
AAATACCCAG TAGATATTTC TCATACTATA CTGGAGTGGA ATTATTTTAA TTGTTCAATC   1080
TCTAAGAACA GCAGTAAAAT GGATCTTTTT ACATCCAACA ACACATTGGA ATGGACAGCA   1140
CTGCACAACT ATGATATGGC CATGAGTGAT GAAGGTTACA ATTTGTATAA TGCTGTTTAT   1200
GTTGCGGCCC ACACCTACCA TGAACACATT CTTCAACAAG TAGAGTCTCA GAAAAAGGTA   1260
GAACACAACA GATATTTCAC TGTTTGTCAG CAGATA                             1296
```

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1521 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

```
ATGAAGAAGC TCTGTGCTTT CACTATTTCA TTTTTGTCTC TGAAGTTTTC TCTCATCTTG     60
TGCTGTTTGA CTGAAGCAAG TTGCTTTTGG AGGATAAAGA ATAGTGAAGA TAGTGATGGA    120
GATTTGCAAA GAGAATGTCA TTTTTACCTT TGGGTAATTG ATAAACCTAT TGAAGATAAT    180
TTTTATAATT CAGTTTTAAA TTTTAGAATA TCAGCAAGTG AATATGAGTT TCTTCTGGTA    240
ATGTTTTTTG CTACTGATGA GATCAACAAG AATCCTTATC TTTTACCCAA CATAACTTTG    300
ATATTCAGCA TCGTTGGTGG TCACTGTCAT GATTTATTGA GAGGTCTGGA TCAATCATAT    360
ACACAAATAA ATGGACGTGT GAATTTTGTT AATTATTTCT GTTATTTAGA TGATTCATGT    420
AACATAGGCC TTACAGGACC ATCATGGAAA AAATCCTTAA AACTGGCAAT GGATTCTTCA    480
ATACCAATGG TTTTCTTTGG ACCATTTAAT CCTAACCTAC GCGACCATGA CCGGCTGCCC    540
CATGTCCATC AGGTAGCCCC CAAGGACACA CATTTATCCC ATGGCATGGT CTCCTTGATG    600
TTTCATTTTA GATGGACTTG GATAGGACTG GTCATCTCAG ATGATGACCA GGGTATTCAG    660
TTTCTCTCAG ATTTAAGAGA AGAAAGCCAA AGGCATGGGA TCTGTTTAGC TTTTGTTAAT    720
ATGATCCCAG AAAACATGCA GATATACATG ACAAGGGCTA CAATATATGA TAAACAAATT    780
ATGACATCTT CAGCAAAGGT TGTTATCATT TATGGTGAAA TGAACTCTAC TCTAGAAGTA    840
AGCTTCAGAA GATGGGAAGA TTTAGGTGCT CGGAGAATCT GGATCACAAC CTCACAATGG    900
GATATCATAT TAAATAAAAA AGAATTCACT CTTAATCTCT TCCATGGCCC TATCACTTTT    960
GCACACCACA AAGTTGAGAT TCCTAAATTA AGGAATTTTA TGCAAACAAT GAACACTGCC   1020
AAATACCCAG TAGATATTTC TCATACTATA CTGGAGTGGA ATTATTTTAA TTGTTCAATC   1080
TCTAAGAACA GCAGTAAAAT GGATCTTTTT ACATCCAACA ACACATTGGA ATGGACAGCA   1140
CTGCACAACT ATGATATGGC CATGAGTGAT GAAGGTTACA ATTTGTATAA TGCTGTTTAT   1200
GTTGCGGCCC ACACCTACCA TGAACACATT CTTCAACAAG TAGAGTCTCA GAAAAAGGTA   1260
GAACACAACA GATATTTCAC TGTTTGTCAG CAGGTATCTT CCTTGATGAA AACCAGGGTA   1320
TTTACGAACC CGGTTGGAGA ACTGGTGAAC ATGAAGCATA GGGAAAATCA GTGTACAGAG   1380
TATGATATTT TCATCATTTG GAATTTTCCA CAAGGCCTTG GATTAAAATT GAAAATAGGA   1440
AGCTATATAC CTTGTTTTCC AAAGAGTCAA CAACTTCATA TATCTGATGA TTTGGAATGG   1500
GCCATGGGAG GAACATCAAT A                                             1521
```

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 933 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

```
ATGAAGCAGC TCTGCACTTT CACTATTTCA TTGTTGTTTC TGAAGTTTTC TCTCATCTTG     60
TGCTGTTGGA GTGAACCAAG CTGCTTTTGG AGGATAAAGA AGAGTGAAGA TAATGATGGA    120
GATTTACAAA GGGAGTGTCA TTTTTACCTT TGGAAAACTG ATGAACCTAT TGAAGATAGT    180
TTTTATAATT ATGATTTAAG TTTTAGAATT GCAGGAAGTG AATATGAGCT TCTTCTGGTA    240
ATGTTTTTTG CTACTGATGA GATCAACAAG AATCCTTATC TTTTACCCAA CATGAGTTTG    300
ATGTTCTCCA TCATTGGTGG AAACTGTCAT GATTTATTGA GAAGTCTGGA TCAAGAATAT    360
GCACAAATAG ATGGACATAT GAATTTTGTT AATTATTTCT GTTATTTAGA TGATTCATGT    420
GCCACAGGCC TTACAGGACC ATCATGGAAA ACATCCTTAA AACTGGCAAT GCATTCTTCA    480
ATGCCACTGG TTTTCTTTGG ACCATTTAAT CCTAACCTAC GCGACCATGA CCGGCTGCCC    540
CATGTCCATC AGGTAGCCCC CAAGGACACA CATTTGTCCC ATGGCATGGT CTCCTTGATG    600
TTTCATTTTA GGTGGACTTG GATAGGACTG GTCATCTCAG ATGATGATCA GGGTATTCAG    660
TTTCTCTCAG ATTTAAGAGA AGAAAGCCAA AGGCATGGGA TCTGTTTGGC TTTTGTTAAT    720
ATGATCCCAG AAAACATGCA GATATACATG ACAAGGGCTA CAATATATGA TACACAAATT    780
ATGACATCTT CAGCAAAGGT TGTTATCATT TATGGTGACA TGAACTCTAC TCTAGAAGCA    840
```

AGCTTTAGAA GATGGGAAGA GTTAGGTGCT CGGAGAATCT GGATCACAAC CACACAATGG 900
GATGTCATCA CAAATAAAAA AAGACTTCAC CCT 933

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1236 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

GCAAGTTGCT TTTGGCGGAT AAAGAATAGT GAAGATAATG ATGGAGATTT GCAAAGGGAA 60
TGTCATTTTT ACCTTGGGGC AGTTGATAAA CCAATTGAAG ATAATTTTTA TAATTCACTT 120
TTAAAGTTTA GAATTGCAGC AAGTGAATAT GAGTTTCTTC TGGTAATGTT TTTTGCTACT 180
GATGAGATCA ACAAGAATCC TTATCTTTTA CCCAACATAA CTTTGATGTT CTCCATCATT 240
GGTGGAAACT GTCATGATTT ATTGAGAGGT TTGGATCAAG CATATACACA AATAAATGGA 300
CATATGAATT TTGTTAATTA TTTCTGTTAT TTAGATGATT CATGTGCCAT AGGTCTTACA 360
GGACCATCAT GGAAAACATC CTTAAAACTG GCAATGCATT CTTCAATGCC ACTGGTTTTC 420
TTTGGATCAT TTAATCCTAA CCTACATGAC CATGACCGGC TGCACCATGT CCATCAAGTA 480
GCCACCAAGG ACACACATTT GTCCCATGGC ATTGTCTCCT TGATGTTTCA TTTTAGATGG 540
ACTTGGATAG GACTGGTCAT CTCAGATGAT GACAAGGGTA TTCAGTTTCT CTCAGATTTA 600
AGAGAAGAAA GCCAAAGGCA TGGGATCTGT TTAGCTTTTG TTAATATGAT CCCAGAAAAC 660
ATGCAGATAT ACATGACAAG GGCTACAATA TATGATAAAC AAATTATGAC GTCTTTAGCA 720
AAAGTTGTTA TCATTTATGG TGAAATGAAC TCTACACTAG AAGTAAGCTT TAGAAGATGG 780
GAAAATTTAG GTGCTCGGAG AATCTGGATC ACAACCTCAC AATGGGATGT CATCACAAAT 840
AAAAAAGAAT TCACCCTTAA TCTCTTCCAT GGGACTATTA CTTTTGCACA CCGCAGATTT 900
GAGATTCCTA AATTTAAAAA ATTTATGCAA ACAATGAACA CTGCCAAATA CCCAGTAGAT 960
ATTTCTCATA CTATATTGGA GTGGAATTAT TTTAATTGTT CAATCTCTAA GAACAGCAGT 1020
AAAATGGATC ATATTACATT CAACAACACA TTGGAATGGA CAGCACTGCA CAACTATGAT 1080
ATGGTGATGA GTGATGAAGG TTACAATTTG TATAATGCTG TTTATGCTGT GGCCCACACC 1140
TACCATGAAC ATATTTTTCA ACAAGTAGAG TCTCAGAAAA AGGCAAAACC CAAAAGATTT 1200
TTCACTGTTT GTCAGCAGCA GATATGGAAC AGTGTG 1236

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2412 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

ATGTTCATTT TCATGGAAGT CTTCTTCCTC CTTAATATTA CACTTCTCAT GGCCAATTTC 60
ATTGATCCCA GGTGCTTTTG GAGAATAAAT TTGGATGAAA TAATGGATGA ATATTTGGGA 120
TTATCTTGTG CTTTCATCCT GGCAGCAGTT CAGACACCCA TTGAAAATGA TTATTTCAAC 180
AAGACTCTTA ATGTTCTAAA AACAACTAAA AACCACAAAT ATGCTTTGGC ATTGGTGTTT 240
GCAATGGATG AAATCAACAG AAATCCTGAT CTTTTACCAA ATATGTCTTT GATTATAAGA 300
TACACTTTGG GCCGTTGTGA TGGAAAAACT GTAATACCTA CACCATATTT ATTTCGTAAA 360
AAAAAAGAAA GCCCTATCCC TAATTATTTC TGTAATGAAG AGACTATGTG TTCCTATCTG 420
CTTACAGGAC CCCATTGGGA GGTATCTTTA GGTTTCTGGA AGCACATGAA CAGCTTCTTA 480
TCTCCACGTA TCCTTCAGCT TACCTATGGA CCTTTCCACT CCATCTTCAG TGATGATGAA 540
CAATATCCCT ATCTCTATCA GATGGCCCCA AAGGACACAT CTCTAGCATT GGCAATGGTC 600
TCCTTCATAC TTTACTTTAG CTGGAACTGG ATTGGCCTTG TCATTCCAGA TGATGACCAA 660
GGAAACCAAT TTCTTTTAGA GTTGAAGAAA CAGAGTGAAA ACAAGGAAAT TTGCTTTGCC 720
TTTGTGAAAA TGATCTCTGT TGATGATGTT TCATTTCCAC AAAATACTGA AATGTACTAC 780
AACCAAATTG TGATGTCATC CACAAATGTT ATTATCATTT ATGGAGAAAC ATACAATTTC 840
ATTGATTTGA TCTTCAGAAT GTGGGAACCT CCCATTTTAC AGAGAATATG GATCACCACA 900
AAACAATTGA ATTTCCCTAC CAGGAAAAAA GACATAAGTC ATGGCACATT CTATGGATCA 960

CTTACTTTTC TACCCCACCA TGGTGTGATT TCTGGTTTTA AAAATTTTGT ACAGACATGG 1020
TTCCATCTCA GAAACACAGA TTTATATCTA GTAATGCAAG AGTGGAAATA CTTTAACTAT 1080
GAAGACTCAG CATCTACCTG TAAAATACTG AAGAACAATT CATCTAATGC CTCATTTGAT 1140
TGGCTAATGG AACAGAAGTT TGACATGACC TTTAGTGAGA ATAGTCATAA CATATACAAT 1200
GCTGTGCATG CCATAGCCCA TGCCCTCCAT GAGATGAATC TGCAACAGGC TGATAATCAG 1260
GCAATAGACA ATGGGAAAAA GGAGCCCAGT TCCTCCCACT GCTTGAAGGT AAACTCCTTT 1320
CTAAGAAGGA TTTACTTCAC TAATCCTCCT GGGGACAAAG TGTTTATGAA GCAAAGAGTA 1380
ATAATGCACG ATGAATATGA CATTGTTCAC TTTGTGAATC TCTCACAACA CCTTGGGATT 1440
AAGATGAAGT TAGGAAAGTT CAGCCCATAT TTACCACATG GTCGACACTC TCACTTATAT 1500
GTAGACAGGA TTGAGTTGGC CACAGGAAGA AGAAAGATGC CATCCTCTGT GTGCAGTGCT 1560
GATTGTAGTC CTGGATTCAG AAGATTATGG AAGGAGGGAA TGGCAGCCTG CTGTTTTGTT 1620
TGCAGCCCCT GCCCTGAAAA TGAAATTTCT AATGAGACAA CTGTGGTACT TTGTGTCTTT 1680
GTGAAGCATC ATGACACTCC TATTGTGAAG GCCAATAACA GAAGCCTCAG CTACCTATTA 1740
CTCATGTCAC TCATGTCCTG TTTTCTGTGC TCCTTTTTCT TCATTGGCCT TCCAAACAGA 1800
GCCATCTGTG TCTTACAGCA AATCACATTT GGAATTGTAT TCACTATGGC TGTTTCCACA 1860
GTTCTGGCCA AAACAGTCAC TGTGGTTCTG GCTTTCAAAG TCACAGACCC AGGAAGAAGA 1920
TTGAGAAACT TCCTGGTATC AGGAACACCC AACTACATTA TTCCCATATG TTCCCTACTC 1980
CAATGTGTTC TGTGTGCAAT CTGGCTAGCA GTTTCTCCTC CCTTTGTTGA TATTGATGAA 2040
CACACTCTCC ATGGCCACAT CATCATTGTG TGCAACAAGG GCTCAGTTAC TGCATTCTAC 2100
TGTATCCTAG GATACTTGGC CTGCCTGGCA CTTGGAAACT TCTCTGTGGC TTTCTTGGCC 2160
AAGAATCTGC CTGACACATT CAATGAAGCC AAGTTCTTGA CCTTCAGCAT GCTAGTGTTC 2220
TGTAGTGTCT GGGTCACCTT CCTCCCTGTC TACCATAGCA CCAAGGGCAA ACACATGGTT 2280
GCTGTGGAGA TCTTCTCCAT CTTGGCATCC AGTGCTGGGA TCCTTGGATG TATATTTGTA 2340
CCCAAGATTT ATATCATTTT AATGAGACCA GAGAGAAATT CGACCCAAAA GATCAGGGAA 2400
AAATCATATT TC 2412

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 381 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

ATGTTCATTT TCATGGGAGT CTTCTTCCTC CTTAATATTA CACTTCTCAT GGCCAATTTC 60
ATTAATCCCA GGTGCTTTTG GAGAATAAAT TTGGATGAAA TAACGGATGA ATATTTGGGA 120
TTATCTTGTA CTTTCATCCT GGCGGCAGTT CAGACACCCA CTGAAAAAGA TTATTTCAAC 180
AAGACTCTTA ATGTTCTAAA AACAACTAAA AACCACAAAT ATGCTTTGGC ATTGGTGTTT 240
GCAATGGATG AAATCAACAG AAATCCTGAT CTTTTACCAA ATATGTCTTT GATTATAAGA 300
TACACTTTGG GCCTTTGTGA TGGAAAAACT GTAACACCTA CACCATATTT ATTTCATAAA 360
AAAAAAACAA AGCCCTATCC C 381

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 228 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

ATGAAAAACC TGTGTGTTTT CACTCTTTCC TTTTTCCTCC TGGAGTTTTC TCTGATCTTG 60
TGCCATTTGA CTGAACCCAT TTGCTTTTGG AGGATAAATA ATAATGAAGA TAATGATGGA 120
GATTTGAGAA GTGACTGTGG TTTTTTCCTT GCAGCAGTTG AGGGACCTAC TGACGACTCT 180
TATAATATCT CTGATCTTAG GTTTTCTTTG GACCATTTAA TCCTAAGC 228

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1644 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

```
ATGTTAGAAT TGGCCCATGG CACTCTGACT TTCTCACCCC ATCATGGGGA GATTTCTGAT     60
TTCACAAATT TTATGCAGGA AGTCACCCCT ATCAAGTACC CAGAAGACAT TTTTCTTCAC    120
ATCTTGTGGA ACCAGTATTT CAATTGTCCA CTTTTGCATT CTGAGTGTAA AATCTTTGAA    180
AACTGTATAC CCAATGCCTC TTTGGAATTG TTGCCAGGGG GTGTTTTTGA GCTGGTCATG    240
ACTGAAGAGA GTTACAATGT GTACAATGCT GTGTATGCAG TGGCCCACAG TCTCCATGAG    300
AAGGCTCTCC ATCAAGTAGA AATTCAACCA CAGGATAATA AAGATAGGAC TATATTATTT    360
CCTTGGCAGC TTCACCCTTT TCTGAAGAAC ATTCAGCTGA TAAATTCTGT TGGTGATCGT    420
GTGATTCTGG ACTGGAAAAA GAAGACGGAT ACAGAGTATG ATATTTCCAA TATTTGGAAT    480
TTCCCAACAG GTCTTTCCTT ATTAGTGAAA GTGGGTACAT TTGCTCCAAG TGCTCCCAAG    540
GGGGAACAAC TTTCGATATC TGAACACACA ATTAACTGGC CCATAGGATT TACAGAGATT    600
CCAAAGTCTG TATGCAGTGA GAGCTGCAGT CCTGGACACA GGAAAGTCAT CCTGGAGAGC    660
AAGCCTGCCT GTTGCTTTGA CTGCACTCCT TGCCCAGATA AAGAGATTTC CAACGAGACA    720
GATGTGGGTC AGTGTGTGAA GTGTCCTGAA TCTCATTATG CAAATACAGA GAAGAGTCAC    780
TGCCTGAAGA AGACTATGAC CTTTCTGGAT TATAATGATT CCTTGGGGAC GGGACTCACA    840
CTCATGTCTC TGGGATTCTT TGTTGTCACA GGTCTTGTTA TTGGGGTTTT TATAATCCAC    900
AGAAACACTC CAATTGTGAA GGCCAATAAT AGATCTCTCA GTTATATCCT GCTCATCACT    960
CTCACTCTCT GTTTCCTTTG TCCCTTGCTC TTCATTGGGC TTCCAAACAC AGCCACATGT   1020
ATCCTACAGC AGAACTTGTT TGGACTTCTC TTCACTGTGG CTCTATCCAC AGTGTTGGCC   1080
AAAACTATCA CTGTAGTTAT GGCATTCAAG ATTACTGCTC CAGGAAGAAA GACAAGATGG   1140
TTGCTGATAT TAAGAGCCCC TCAGTTCATC ATTCCACTTT GTGCCCTGAT GCAAATCCTT   1200
TTCTCTGGGA TATGGCTGGG AACATCTCCT CCATTTGTTG ACATGGATGC TCACTCTGAA   1260
CATGGGCACA TCATCATTCT ATGCAACAAG GGCTCAGCTA TTGGCTTCTA CTGTACTCTG   1320
GCCTACCTGG GAGTCATGGC CTTTGGTAGT TACCTCTTGG CTTTCATGTC CAGGAATCTT   1380
CCTGACACAT TTAATGAATC CAAGGCCCTG GCTTTCAGCA TGCTGATGTT CTGCAGTGTC   1440
TGGGTCACAT TCCTCCCTGT CTACCACAGC ACCACTGGGA AGGTCAGGGT GGCTATGGAA   1500
ATGTTTTCTA TCTTGGCTTC CAGTGCAAGC ATTCTAACCC TAATCTTTGT CCCTAAGTGC   1560
TACATTGTTT TGTTCAGACC AGAGAGGAAC ATACTTCCTC TAAACAGAGA AAAAAGACAG   1620
CATAGGAGTA AAAATTCTGA AACA                                          1644
```

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2304 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

```
ATGGAGGAAA TCAACAGGAA CCCTGATCTT TTACCAAATA TGTCTTTGGT TATAAAACAT     60
ACTTTGAGCT ATTGTGATGG AAATACTGCA GACCATATAT TTAAAGAAAA ATTTTATAAG    120
CCTTTACCTA ATTATGTCTG TAATGAAGAG ACTATGTGTT CATTTATGCT TATAGGGCTG    180
AATTGGGTAT TGTCTCTAAC ACTTTTTAAA GACTTGGACA TCTTCTCATT TCCACGTTTC    240
CTTCAAATTT CCTATGGACC TTTCCATTCC ATCTTCAGTG ATAATGAACA ATTTCCATAT    300
CTCTATCAGA TGACCCCAAA GGACACATCA CTAGCATTGG CAATTGTCTC CTTCTTACTT    360
TACTTCAATT GGAACTGGGT TGGGCTTGTC ATCTCTGATA ATGATGAAGG CAATCAATTT    420
CTCTCAGAGT TGAAAAAAGA GACCCAAAAC AAGGAAATTT GCTTTGCCTT TGTTAACATG    480
ATGTCAATCC ATGAGCATTC ATCTTATCAA AAAACTGAAA TGTACTACAA TCAAATAGTG    540
ATGTCATCAA CAAATATTAT TATCATTTAT GGGAAAACAA ACAGTATCAT TGAATTGAGC    600
TTCAGAATGT GGGTATCTCC AGTTATACAG AGGATTTGGG TCACAAACTC AGAGTTGGAT    660
TTCCCGACAA GTATGAGAGA CTTCACTCAT GGCACATTCT ATGGGACTCT GACATTTCTA    720
CACCACCATG GTGAGATTTC TGGATTTACA AATTTTTTCG AGACATGGGA CCATCTCAGA    780
AGCAGAGATT TAAATCTATT AATACCAGAG TGGAAGTACT TTAGCTATGA TGCCTCAGGA    840
```

```
TCTAACTGTA AAATATTGAG GAACTATTCA TCCAATGCCT CATTGGAATG GATAACAGAA    900
CAGAAGTTTC ACATGGCCTT TAATGATTAT AGTCATAGTA TATATAATGC TGTGTATGCC    960
ATGGCCCATG CCCTCCATGA GACTAATCTG CAAGAGGTTG ATAATAAGGA AATAAGAAAT   1020
GGGAAAGGAG CAAGTACTCA CTGCTTGAAG GTAAACTCAT TTCTCAGAAA GACCCACTTT   1080
ACTAATTCTC ATGGAGAGAG AGTGATTATG AAACAGAGAG TGAGAGTACA GGAAGACTAT   1140
GACATTGTTC ACATTCAGAA TTTCTCACAA CACCTTCGGA TTAAGATGAA GATAGGAAAG   1200
TTCAGCCCAT ATTTTACACA TGGTGGACCC TTTCACTTAT ATGAAGACAT GATTCAGTTG   1260
GCCACAGGAA GTAGAAAGAT GCCGTCCTCT GTGTGCAGTG CAGATTGTAG TCCTGGATTC   1320
AGAAAATCCT GGAAGGAGGG AATGGCCCCC TGCTGTTTTA TTTGCAGCCT GTGCCCTGAA   1380
AATGAAATTT CTAATGAGAC AAATATGGAT CAATGTGTGA ATTGTCCAGA ATACCAATAT   1440
GCCAACACAG AAAAGAACAA ATGCATTCAG AAAGACGTGA TTTTTCTAAG CTATGAAGAC   1500
CCCTTGGGAA TGGCTCTTGC CTTAATTGCC TTCTGTTTGT CTGCATTCAC AGCTGTGGTA   1560
CTTTGGGTCT TTGTGAAGCA CCATGACACT CCTATTGTGA AGGCCAATAA CAGAATCCTC   1620
AGCTACATAT TAATCATGTC ACTAATGTTC TGTTTTCTCT GCTCCTTTTT CTTCATTGGC   1680
CATCCTAACA GAGGTACCTG TATCTTACAG CAAATCACAT TTGGCATTGT ATTCACTGTG   1740
GCTGTTTCCA CAGTTCTGGC CAAAACAATC ACTGTCATTC TTGCTTTCAA ACTCAGAGAC   1800
CCAGGGAGAA GTTTAAGAAA CTTCCTGGTA TCTGGTGCAC CCAACTACAT TATTCCTATA   1860
TGTTCCTTAT TGCAATGTAT TCTGTGTGCA ATTTGGCTAG CAGTTTCTCC TCCTTTTGTT   1920
GATATTGATG AACATTCTGA GCATGGCCAC ATCATGATTG TGTGCAACAA GGGCTCCATT   1980
ATGGCATTCT ACTGTGTCCT AGGATACTTG GCCTGCCTGG CGCTTGGAAG CTTCACTACA   2040
GCTTTCTTGG CAAAGAATCT GCCAGACACA TTCAACGAAG CCAAGTTCTT GACCTTCAGC   2100
ATGCTAGTGT TCTGCAGTGT CTGGGTCACC TTTCTCCCTG TGTACCATAG CACAAGGGGC   2160
AGGGTCATGG TTGCTGTTGA GATCTTCTCT ATCTTGGCAT CCAGTGCAGG GATGTTTGGA   2220
TGCATCTTTG CACCCAAAAT CTACATCATA TTAATGAAAC CAGAAAGAAA TTCTATACAA   2280
AAGTTCAGGG AGAAATCATA TTTC                                          2304
```

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2001 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

```
ATGGCTCCTA AGGACACATC TCTGGCACTG GCCATGGTTT CTTTGTTTGT CCATTTCAGC     60
TGGAACTGGG TAGGAGCTGT TGTTTCAGAT GATGACCCAG GTTATGAATT TATCTTGGAA    120
TTGAGAAGAG AAATGCAAAG GAACAATTTT TGTTTAGCAT TTGTGAGTAT CATTGTTAGT    180
GATGACAATT TATTTCTGAA AAGGTATAAT ATCTATTACA ACCAGATCAA GATGTCATCA    240
GCAAAAGTTG TTATCATTTA TGGAGACAAA GACTCTCCTC TACAGGTGAA CTTTAGACTA    300
TGGAATTTAT TTGATATCCA AAGAATCTGG GTCACTACTT CACAGTGGGA TATGATCATA    360
AATAATGGAA AATTCCTCCT TAATTCCTTC TATGGGACTC TCAGTTTTTC ACATCACTAT    420
TCTGAATTAT CTGGTTTTAA AACATTTATC CAGACAGCAT ACCCTTCAAA CTACAGTGAT    480
GACTTTTCTC TTGGTATATT ATGGTGGGTG TATTTTAATT GTTCTTTGTC ATTATCTGAA    540
TGTAAGAATC TGCAAAATTG TCCAAAGGAA AACATATTTA GATGGTTATA CAGGCACCAT    600
TTTGAAATGT CTTTGAGTGA TACTACTTAT GACCTATATA ATTCTATGTA TGCTGTGGCT    660
TACACACTCC AACAGATGCT TCTGAAACAA GCAGATACAT GGCAAATAGA TGATGGAAAA    720
GAACCAGAAT TTGACTCTTG GCAGATGCTC TCTTTCCTGA GAAATATCCA ATTTATAAAC    780
CCTGTTGGTG ACAAAGTGAA CCTGAATCAT GAAGAAAAAC TGGATACAAA GTATGAGATT    840
CACCAGACTT TGACTTTTTT GCCAAATCCT GTATTTAAGC TGAAAATAGG AACATTTTCC    900
CAAAACTTAT CACATGGTCG ACAATTATAT ATGTTGAAAG AAATGATAGA GTGGAACACA    960
GGCCACCAAC AGTCTCCAAC CTCAGTTTGC AGTATTCCTT GTAGTCCAGG ATTCAGAAAA   1020
TCCCCTCAGC TGGGAAAGCC TGTTTGCTGT TTTGATTGTA CACCCTGCCC AGAAAATGAA   1080
ATTTCCAACA TGACAAACAT GAATCAATGT ATCAAGTGTC TAAATGATCA GTATGCCAAT   1140
CCTGGAGGAA CTCGCTGCCT CAAAAAAGTT ATTGTATTCC TGGGTTATGA AGATCCATTG   1200
GGAATGTCTC TGGCTATCTT GGCTCTGTGC TTCTCTGCTC TCACAGCTTT TGTACTTAGT   1260
ATCTTTTTGA AGCACCAAGA AACACCCACT GTCAAGGCCA ATAATAGAAC TCTCAGCTAT   1320
GTTCTACTCA TCTCCCTCAT CTCTTGTTTT CTCTGCTCCT TGCTCTTCAT TGGTCATCCC   1380
AGCTTTACCA CATGTATCAT GCAGCAGACC ACATTTGCTG TTGTGTTCAC TGTAGCTGCA   1440
TCTACTGTCT TGGCCAAAAC AATTATTGTA ATATTGGCCT TCAAGGTTAC TAATACAAGT   1500
AGAAAAATGA GGTGGCTGCT GGTATCAGGG GCACCTAAAT TCATCATTCC AATTTGCACA   1560
ATGATTCAAC TGATTCTCTG TGGAATTTGG CTGGGTACTT CTCCTCCATT TGTTGATGCT   1620
```

GATGGACATG TTGAAAAAGG CCACATTTTG ATTTTCTGTA ACAAAGGTTC AATTCTTGCT 1680
TTCTATTGTG TCCTGGGATA CTTAGTCTCC ATTGCCATTG CAAGTTTCAC CCTTGCATTC 1740
TTCGCCAGAA ATCTGCCCGA CACATTCAAT GAAGCCAAGT TCCTAACATT CAGTATGCTA 1800
GTATTTTGCA GTGTCTGGGT CACCTTTCTT CCTGTCTATC ATAGCACCAA GGGCAAGTCT 1860
ATGGTGGCTG TGGAAGTTTT CTGTATATTG GCCTCTAGTG CAGGGCTGCT TTTTTGCATC 1920
TTTGCACCAA AGTGCTTCAT TATTTTGTTA AGACCTGAGA AAAAATCTTT TCAGAAGTTT 1980
CAGAATATAC ATTCTAAAAT T 2001

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2598 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

ATGTCCAGGC TCAGAGCAGG AAAAAATATG CTCACCTTCA TTTTACTCTT CTTTCTCCTG 60
AACATTCCAC TTTTTGTGCC TAGTTTTATT TATCCCAGGT GCTTTTGGAG TATGAAGAAG 120
AATGAATATC AGGATAGAAA CCTGGGAACA GGTTGTATGT TCTTTATTCT AGCAGTGCAA 180
CAGCCTATGG AAAAAGAGTA TTTCAGTCAT ATTTCGAATA TACAAACACC TACTGAAAAC 240
CAAAAGTATC CTCTCACCTT GGCTTTTTCC ATGAATGAAA TCAACAACAA CCCTGATCTT 300
TTGCCAAATA TGTCTTTAGC ATTTACATTC TCAGAATATA GTTGTTATTT GGAATCCCAC 360
CACAAAAGAT TATTTAATTT TTCTTTAAAA AATCATGAAA TTCTCCCTAA TTTTATCTGT 420
ACAAAAGACA TCAAGTGTGG AGTGGTACTT ACCGGACTTA GTTTGGTAAC AACTGTGACA 480
CTTCATATAA TCCTAAACAA TTTCATATTT CAGCAGTTCC GTCAGCTTAC TTATGGACAC 540
TTTCATCCTG CTCTGTGTGA TCATGAAAAT TTTCCTCATC TATATCAGAT GGCCTCTGAT 600
GATACATCTC TAGCCCTTGC TCTCGTCTCC TTCATAATTC ATTTCAGTTG GAACTGGATA 660
GGGTTGGCCA TCTCAGACAA TGATCAAGGC ATACATTTTC TCTCTTATTT GAGAAGAGAG 720
ATGGAAAAAA ATACAGTCTG CTTTGCCTTT GTCAACATTA TTCCAGTCAA TATGAATTTA 780
TACATGTCAA GAGCTGAAGT GTATTACAGC CAAGTTATGA CATCATCCGC AAATGTTGTT 840
ATCATTTATG GTGATACAGG GAATACGTTA GCTGTGAGCT TTAGAATGTG GGACTCTCTA 900
GGTATACAGA GACTATGGGT CACCACCTCA CAGTGGGATG TCACTCCTTT TAAGAAAGAC 960
TTCACATTTG ATAATGGATA TGGAACTTTT GGTTTTGGAC ACCGCCACAG TGAGATTTCT 1020
GGTTTTAAAT ATTTTGTTCA GACATTGAAC CCTTTCAAAT ACTCAGATGA ATATTTGGTA 1080
AAGCTGGAAT GGATGTATGT TAATTGTAAA ATCTTAGAAT ATAACTGTAA GTCACTGAAG 1140
AACTGCTCCT TTAATCACTC ATTGGAATGG CTAATGACAC ATACTTTTGA CATGGCCATT 1200
ATTGAAGGGA GTTATGAAAT ATACAATGCT GTGTATGCTT TTGCCCATGC ACTCCATGAG 1260
ATGACTCTTC AAAATGTTGA TAATGTTCTC CTTCCCAATT ATGAAGAACA AAATTATAAT 1320
TGCAAGATGG TTTATTCCTT TCTGAGCAAG ACTCAATTCA CAAATCCTGT TGGAGACACT 1380
GTGAATATGA ATCAAAGAAA CAAACTGAAG GAAGAGTACG ACATTTTCTA CAATTGGAAT 1440
TTTCCACAGG GACTTGGATT TAAAGTGAAA ATAGGAATAT TTAGTCCATA TTTTCCAAAA 1500
GGTCAACAGC TTCATTTATC TGAAAATCTG ATAGAGTGGT CCACAGGACG TATACAGATG 1560
CCAACCTCTG TGTGCAGTGC CGATTGTGGT CCTGGATTTA GGAAAGTCTG GAAGAATGGA 1620
ATGCCAGCCT GTTGTTTTGA CTGCAGTCCC TGCCCAGAAA ATGAAATTTC TAATGAGACA 1680
AATGTGGAAT TGTGTGTCCA GTGTCCAGAG GACCAATATG CTAACCAAGA GCAGAATCAC 1740
TGCATTCACA AAGCTCGTAT CTTTCTCTCT TATGATGAAC CCTTGGGGAT GGCTCTTTCC 1800
TTAATGGCCT TATGCCTCGC TGCACTCACA GTTGTGGTTC TTGGAGTCTT TGTGAAACAT 1860
CACAGAACTC CCATAGTTAA GGCCAATAAC TGCACTCTCA CCTACATCTT GCTCATCGCA 1920
CTCATCTTTT GTTTCCTCTG CCCCTTGTTC TTCATTGGCC ATCCAAACTC AGCTACCTGC 1980
ATCCTTCAGC AAATCACATT TGGAGTTGTG TTCACTGTGG CTATTTCCAC TGTGTTGGCC 2040
AAAACAACCA CTGTCATTCT GGCTTTCAGA GTCACAGCCC CTCATAGAAT GATGAAGTAC 2100
TTTCTTGTTT CAAGGGCATC TAACTACATC ATTCCCATTT GTACTCTCAT TCAAATTATT 2160
GTATGTGCCA TCTGGCTAGG AGCTTCTCCT CCTTCTGTTG ATATTGATGC ACAGTCTGAG 2220
CATGGTCACA TCATCATTGC TTGCAACAAG GGTTCAGTCA CTGCTTTTTA CTGTGTCCTG 2280
GGATATCTGG CCTGCCTGGC CTTTGTGAGC TTCACCCTGG CTTTCCTTTC CAGAAACCTG 2340
CCTGTCACCT TCAATGAAGC CAAGTCCATG ACATTCAGCA TGCTGGTGTT CTGCAGTGTC 2400
TGGGTCACTT TCCTACCTGT TTACCATGGC ACCAAAGGCA AGGTTATGGT GGCTGTTGAG 2460
ATCTTTTCCA CCTTGGCTTC TAGTGCAGGA ATGTTGGGAT GCATTTTTGC TCCAAAATGC 2520
TACACAATAC TGTTTAGACC AGACAGAAAT TCTCTTCAAA TGATCAGGGA GAAGTCATCT 2580
TCTCATACTC ACATTTTA 2598

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2337 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

ATGAGGTTTG CCATTGAGGA AATCAACAGC AATCCCCATC TTTTACCAAA CACATCCCTG 60
GGATTTGAGA TCAATAATGT CCCACACGGT CAGAGGTACA CTCTGGTCAA ACTTTTTAGC 120
TCACTTTCAG GGTCTAATTA TGACATTCCT AACTACATAA GTGCAAGTGA GAGCAATTCT 180
GCTGCTGTAC TTACAGGACC ATCGTGGACA ATATCTGAAT GCGTAGGGAC ACTCCTGGAT 240
CTTTACAAAT TTCCACAGCT TACTTTTGGG CCTTTTGATA GTCTCCTGAG TGAACAAAGA 300
CGGTTTTCTT CTCTGTACCA AGTGGCCCCC AAAGATACAT TTCTGACGCC TGGCATTGTA 360
TCTTTGATGC TTCATTTCCA CTGGAACTGG GTGGGGTTAT TCATCATAGA TGATGACAAA 420
GGTGCCCAGA CACTGTCAGA CTTGAGAAAT GAGATGGATA AAAATGGAGT CTGCACAGCA 480
TTTGTAGAAA TGATCCCAGT CATCAAGGGT TCATTTTTTA CCAAATCCTG GAAAAATCAT 540
GTGCAGATCC TGGAATCATC ATCAAATGTG ATTATTATTT ATGGGGACTC TGATTCTCTA 600
TTAAGCTTAA TAGTAAATAT TAAGCAGAAG TTGCTCACAT GGAAAGTGTG GGTACTGATC 660
TCACAGTGGG ATGTTTCTAA ATTTGATGAT TATTTCATGG TAGACTCATT GCATGGAGCT 720
CTTATTTTTT CACACCATCG TGAGGAGATT CCTAATTTTA CAGATTTTAT GCAGAAGTAC 780
AACCCTTCCA AGTACCCGGA AGACACTTAT CTTCATGTAT TGTGGCACAT GTACTTCAAT 840
TGCTCATTTG TTAAGAAAGA TTGTAAAATT GTGCACAACT GTTTGCCTAA TGCCTCCCTG 900
GGGTTCTTGC CTGGGAACAT ATTTGACATG GCCATGAGTG AAGAGAGTTA CAATGTATAC 960
AATGCTGTGT ATGCTGTGGC CCACAGTCTG CATGAGATGA TTCTCAACCA AGTACAATTT 1020
CAAACTCATG AAAAAGGAAA AAAGATGGTA TTCTTTCCTT GGCAGCTTCA CCCCTTTCTA 1080
AGGGAAAGAC AACTCATCAA TCAGAATGGA GCGAATGAAG ATCTGGATTG TACCAGGAAG 1140
TCACATGTAG AGTATGACAT TCTCAACTTT TGGAATTTCC CAAAAGGTCT TGGGCTAAAT 1200
GTGAAAGTAG GAACGTTTTC TCCAAGTGCT CCAAAGGAAC AGAAACTGTC CATATCTTCT 1260
AACATGATAC AGTGGGCCAC AGGGTCGACA GAGATTCCAC AGTCTGTATG CAGTGAGAGC 1320
TGTCATCCTG GATTCAGGAA AACCCACCAG GAAGGCAGGG TTGCCTGTTG CTTTGACTGC 1380
ATTCCTTGTC CAGAAAATGA GATCTCCAAT GAGACAGATG TGGATCAGTG TGTGAAGTGT 1440
CCAGAAACTC ACTATGCAAA CATAGAGAAG ATCCACTGCC TACAGAAAAC TGTGACATTT 1500
CTGTACTATG ATGACCCATT GGGGAAGACA CTTTGCTTCA TGTCCCTGGG TTTCTCCTCA 1560
CTCACAGCTG CTGTTCTTGT GGTGTTTCTG AAGAACAGGG ACACCCCCAT TGTCAAGGCC 1620
AATAACCTGG CTCTCAGTTA CACCCTGCTC ATCACTTTGA TGCTCTGTTT TCTCTGTCCC 1680
TTGCTCTTCA TTGGCCGTCC CAGCACAGCC TCCTGTATCC TGCAGCAAAA CATTTTTGGG 1740
CTTCTGTTCA CTGTGGCTCT TTCCACTGTG TTGGCCAAAA CTATCACTGT GGTTATAGCC 1800
TTCAAGATCA CTTCTCCAGG AAGAATTAGA AGATGGCTGC TGATATCAAG GGCCCCTAAT 1860
TTCATTATTC CCTTATGCAC CCTGCTCCAA GTTTTTCTAT CTGGAATTTG GCTGACAACC 1920
TCTCCTCCAT TTATTGATAA AGATGCTCAC TCAGAACATG GACACATCAT CATCATTTGC 1980
AATAAAGGCT CAGCTGTTGC TTTCCATTGC AACCTTGGAT ACCTGGGAGC ACTAGCCCTA 2040
GTGAGCTACT TTATGGCTTT CTTGTCCAGA AACCTACCTG ACACATTCAA TGAAGCCAAG 2100
TTCCTGGCTT TCAGCATGCT GGTGTTCTGC AGTGTCTGGG TCACCTTCCT CCCTGTCTAC 2160
CACAGCACCA AGGGGAAGAA CATGGTGGCT ATGGAAGTCT TCTCTATCTT GGCTTCCAGT 2220
ACATCTCTCC TAGGCATCAT CTTTGCCCCC AAGTGCTACC TCATATTATT AAGACCAGAA 2280
AGGAATTCAC TTAGCTATAT CAGGGACAAA ACATATGCTA AAAGCATAAA ACCTTCT 2337

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1650 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

ATGAAGTTAA GGGATAAAGA CTTGAGCATA ACTTGTTCCT TCATCCTTGA AGCAGTTCAG 60

ATGCCTACGG AAAACGATTA TTTCAACCAG ACTCTGAATA TCCTAAAAAC AACAAAAAAC 120
CACAAATATG CTTTGGCATT GGCCTTTTCA ATTGATGAAA TCAACAGGAA TCCTGATCTT 180
TTACCAAATA TGTCTTTGAT CATAAAATAC CCTTTGGGCC TTTGCGATGG ACAAACTACA 240
TTACCTACAC CCTATTTATT TAATGAAATA TATTTTAGGC CTATCCCTAA TTATTTCTGT 300
AATGAAGAGA CTATGTGTAC ATTTCTACTT ACAGGACCGC ATTGGATAAC ATCTTATAGT 360
TTCTGGATAC ACTTGAACAT CTTCTTATCT CCTAGTATGA ACCCAAAGGA CACATCCCTA 420
GCTTTGGCAA TGGTCTCCTT CTTACTTTAT TTCAAGTGGA ACTGGGTCGG CCTTGTCATC 480
TCAGATGATG ATCAAGGCAA TCAATTTCTC TCTGAGTTGA AAAAAGAGAG CAAAATCAAG 540
GAAATTTGCT TTGCATTTGT GAGCATGCTG GCAATCGATG AGATTTCATT TTATCATAAA 600
ACTGAAATGT ACTACAACCA AATTGTGATG TCATCCACAA ACGTTATTAT CATTTATGGG 660
AAAACAGAGA GTATTATTGA GTTGAGCTTC AGAATGTGGG AATCTCCAGT TATCCAGAGA 720
ATATGGGTCA CCACAAAAGA AATGAATTTC CCTACCAGTA AGAGAGATTT AACTCATGAC 780
ACATTCTATG GGACTCTTAC TTTTCTACAC AGCCATGGGG AGATTTCAGG CTTTAAAAAT 840
TTTGTACAGA CATGGTACCA TCTTAGAATC ACTGATTTGC ATCTAGTAAT GCCAGAGTGG 900
AAATATTTTA ACTATGAAGC CTCAGCATCT AACTGTAAAA TATTGAAGAA CTATTCATCC 960
AGTGCCTCAT TGGAATGGTT AATGGAGCAG ACATTTGACA TGGTCTTTAG TGATGGAAGT 1020
CGGGATATAT ATAATGCTGT AAATGCCATG GCCCATGCAC TCCATGAGAT GAATCTGCAC 1080
CTGGTTGATA ATCAGGCAAT AGACAATGGG AAAGGAGCCA GTTCTCACTG CTTTAAGATA 1140
AACTCCTTTC TCAGAAAGAC CCACTTCACT AATCCTCTTG GGGACAGAGT GATTATGAAA 1200
GAGAGAGAAA TACTGCAAGA AGACTATAAC ATTTTTCACA CTTGGAATTT TTCTCAGCAC 1260
ATTGGTTTTA AGGTGAAGAT AGGAAAGTTC AGCCCATATT TTCCACATGG CAGGCACTTT 1320
CACCTATATG TAGACATGAT TGAGTTGGCT ACAGGAAGTA GAAAGATGCC ATCCTCTGTG 1380
TGCACTGAAG ATTGTAGTCC TGGATACAGA AGATTCTGGA AGGAGGGAAT GGCAGCCTGC 1440
TGTTTTGTTT GCAGTCCCTG CCCTGAAAAT GCAATTTCTA ATGAGACAAA TATGGATCAG 1500
TGTGTGAATT GTCCAGAATA CCAATATGCC AATACAAAGC GGGACAAATG CATTCAGAAA 1560
AATGTGATGT TTCTAAGCTA CAAAGACCCC CTTGGGGATG ACTCTTGCCT TCATAGCCTT 1620
CTTTTTCTCT GCATTAACAG CTGTTGTACT 1650

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2379 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

ATGATAGTAT TCTTTCTCCT CAACATTCCA CTTCTCATGG CAAATTCCGT TGATCCCAGG 60
TGCTTTTGGA AAATAAATTT GAATGAAGTC AAGGATATAG ATTTAGATAC AAGTTGTTAC 120
TTCATCCTTG AGGCAGTTCA GTTGCCTATG GAGAAAGATT ATTTCAACCA GACTCTGAAT 180
GTCCTAAAAA CAACCAAATA CAACAGATAT GCATTGGCAT TAGCCTTTAC AATGGATGAA 240
ATAAACAGGA ATCCTCATAT TTTACCAAAC ATGTCTTTGA TTATAAAACA TACATTGGGC 300
CACTGTGATG GAAATATCCC ACTCCGCTTA CTTAATCAAA TATTTTATAT GCCTTTTCCT 360
AATTATGGCT GTAATGAAGA GACTATGTGT TCATTTATGC TTATGGGACC GAATTTGTGG 420
CCATCTGTAG ATTTTTTCAT TCACTTGAAC ATCTTATTTC CTCATTTCCT TCAGATTTCC 480
TTCGGACCTT TCCATTCCAT TTTCAGTGAT AATGAACAAT TTCCTTATAT CTATCAGATG 540
ACCCCAAAGG ATACATCACT AGCATTGGCA ATGGTCTCTT TCATACTTTA CTTCAACTGG 600
AACTGGGTTG GTCTTGTCCT CTCAGATAAT GATGAAGGCA ATCAATTTCT CACAGAGTTG 660
AAAAAAGAGA CCCACAACAC GGAAATATGC TTTGCCTTTG TGAACATGAT GGCAATCAAT 720
GAGAATTCAT CCATGAAAAA AACTGACATG TACTACAACC AAATTGTGAT GTCAACCGCA 780
AATGTTATTA TCATTTATGG GGAACGACCC AGTATTATTG AACTGTGTTT CAGAACATGG 840
ACATCTCCAG TCATACAGAG GATATGGGTT ACCAAATCAG AGTTGTATTT CCCAACAAGT 900
AAGAGAGACT TAAGTCATGG AACATTCTAT GGAACTCTAG CATTTCAACA ACACCATGAT 960
GTGATTTCTG GATTTAAAAA TTTTGTACAG ACATGGTACC ATCTCAAAAG CATGGATTTA 1020
TATTTATTAA AGCCAGAGTG GGGTTTCTTT GAATATGAAA CCTCAGCATC TTACTGTAAA 1080
ATACTGATGA GTAATTCATC GAATGTCTCA TTGGAATGGC TAATGGAACA GAAGTTTGAC 1140
ATAGCCTTTA ATGACAATAG TCATAGTATA TACAATGCTG TGTACGCCAT GGCCCATGCT 1200
CTCCATGAAA AGAATCTGAA ACAAATTGAT AATCAGGAAA TCAGCTATGG CAAAGGAGCA 1260
AGTACTCACT GCTTGAAGTT ACACTCATTT TTGAGAACGA TCCACTTCAC CAATCCTTTT 1320
GGGGAGAGAG TGATTATGAA AGAGAGAGTA AGAGTGCAGG AAGACTATGA CATTGTTCAC 1380
CTGCAGAACT GCTCACAACA CCTTAGGATT AAGGTGAAGA TAGGGCAGTT CAGCCCATAT 1440
TTTCCACATG GTGGACAATT TCACTTATAT GAAGACATGA TTGATTTGGC CACAGGAAGT 1500

AGAAAGATGC CTTTATCTAT GTGTAGTGCA GATTGTCGTC CTGGATACAG AAAATTCTGG 1560
AAGGAGGGAA TGGCAGCCTG CTGTTTTGTT TGCAGTCCCT GTCCAGACAA TGAAATTTCT 1620
AATGAAACAA CTGTGGTACT TTGGGTCTTT GTGAAGCACC ATGACACTCC TATTGTGAAG 1680
GCCAATAACA GAATCCTCAG CTACATATTA ATCATGTCAC TCATGTTCTG CTTTCTGTGC 1740
TCCTTTTTCT TCATTGGCCA TCCTAACAGA GGTACCTGTA TCTTACAGCA AATCACATTT 1800
GGAATTGTAT TCACTGTGGC TGTTTCCACA GTTCTGGCCA AAACAATCAC TGTGCTTCTG 1860
GCTTTTCAAG TCACAGACAC AGGAAGAAAG TTAAGAAACT TCCTGGTATC GGGGACACCC 1920
AACTACATTA TTCCCATATG TTCCCTGTTG CAATGCACTC TGTGTGCAAT TTGGCTAGCA 1980
GTTTCTCCAC CATTTGTTGA TATCGATGAA CATTCTGAGC ATGGTCACAT CATAATTGTG 2040
TGCAACAAGG GATCTGTTAT GGCATTCTAC TGTGTCCTGG GATATTTGGC CTTCCTGGCC 2100
CTTGGAAGTT TCACGATGGC TTTCTTGGCA AAGAATCTGC CTGACACATT CAATGAAGCC 2160
AAGTTCTTGA CCTTCAGCAT GCTAGTGTTC TGCAGTGTCT GGATCACGTT CCTTCCTGTC 2220
TACCATAGCA CCAAGGGCAG AGTCATGGTT GCTGTTGAAA TTTTCTCCAT TTTGACATCC 2280
AGTGCAGGGA TGCTTGGATG CGTCTTTGCA CCCAAAATTT ACATCATTTT AATGAAACCA 2340
GAGAGAATTC TATCCAAAAG ACAGGAGAAA TCACGTTTC 2379

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2394 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

ATGGTAATAT TCTTCCTTCT CAACATTCCA TTTCTCCTGG CAAATTTCAT GGATCCCAGA 60
TGCTTTTGGA AAATAAATTT GAATGAAATC AAGGATGAAG TCCTTGGGAT GACTTGTTCC 120
TTCATCCTTG AAACAGTTCA GAAGACTATG GACAAAGATT ATTTCAACCA GACTCTGAAT 180
GTCCTAAATA CAACTACAAA CCACAAATAT GCCTTGGCAT TGGCCTTTAC AGTGGATGAA 240
ATCAACAGGA ATCCTGATCT TTTACCAAAT ATGTCTCTGA TTATAAAATA CAATTTGGGT 300
CATTGTGATG GAAAAACTGT AACAACTCTA TCCGATTTAT TTAATCCAAA TAATCATCTC 360
CATTTCCCCA ATTATTTATG TAATGAAGGG ATTATGTGTT TGGTTCTGCT TACAGGACCA 420
CATTGGAGAG CATCTTTATA TCTCTGGATA TCCGTGTATG TCTACCTGTC TCCACATTTC 480
CTTCAGCTTT CCTATGGACC TTTCTACTCC ATCTTCAGTG ATAATGAACA ATATCCTTAT 540
CTCTATCAGA TGGGCCCAAA GGACTCATCA CTAGCATTGG CAATGGTCTC CTTCATAATT 600
TACTTCAAGT GGAACTGGGT TGGGCTATTT ATCTCAGATG ATGATCAAGG CAATCAATTT 660
CTCTCAGAGT TGAAAAAAGA GAGCCAAACC AAGGATATTT GCTTTGCCTT TGTGAACATG 720
ATATCAGTCA GTGATGTTTC ATACTATCAT AAAACTGAAA TGTACTACAA CCAAATTGTG 780
ATGTCATCCA CAAAGGTTAT TATCATTTAT GGGGAAACAA ACAGTATTAT TGAATTGAGC 840
TTCAGAATGT GGTCATCTCC AGTTAAACAG AGAATATGGG TCACCACAAA ACAATTTGAT 900
TGCCCTACCA GTAAGAGAGA CTTAACTCAT GGCACATTCT ATGGGACCCT TACATTTCTA 960
CACCACTATG GTGAGATTTC TGGCTTTAAA AATTTTGTAC AGACACGGTA CAATCTCAGA 1020
AGCACAGATT TATATCTAGT AATGCCAGAG TGGAAATATT TTAACTATGA AGCCTCAGCA 1080
TCTAACTGTA AAATACTGAG AAACTATTTA TCCAATATCT CACTGGAATG GCTAATGGAA 1140
CAGAAATTTG ACATGTCATT TAGTGATTAT AGTCACAACA TATACAATGC TGTATATGCC 1200
ATTGCTCATG CACTCCATGA GAAGAATCTG CAAGAAGTTG AAAATCAGGC AATAAACAAT 1260
GCGAAAGGAG AAAATACTCA CTGCTTGAAG CTAAACTCAT TTCTGAGAAA GACCCACTTC 1320
ACTAATTCTC TTGGGAACAG AGTAATTATG AAACAGAGAG AAGTAGTGCA TGGAGACTAT 1380
AATATTGTTC ACATGTGGAA TTTCTCACAA CGCCTTGGGA TTAAGGTGAA GATAGGACAA 1440
TTCAGCCCAC ATTTTCCACA GGGTCAACAG TTACACTTAT ATGTAGACAT GACTGAGTTG 1500
GCTACAGGAA GTAGAAAGAT GCCATCCTCA GTGTGCAGTG CAGATTGCCA TCCTGGATTC 1560
AGAAGAATCT GGAAGGAGGA AATGGCAGCC TGCTGTTTTG TTTGCAACCC CTGCCCTGAA 1620
AATGAAATTT CTAATGAGAC GATGGTGGTA TTTTGGGTCT TCGTGAAGCA CCATGACACT 1680
CCTATTGTGA AGGCCAATAA CAGAATCCTC AGCTACCTAT TAATCGTGTC ACTCATGTTC 1740
TGTTTTCTGT GCTCCTTTTT CTTCATTGGC TATCCTAACA GAGCAACCTG TATCTTACAG 1800
CAAATCACAT TTGGAATCTT CTTTACTGTG GCTATTTCCA CAGTTCTGGC CAAAACAATC 1860
ACTGTGGTTC TGGCTTTCAA AGTCACAGAC CCAGGAAGAC AATTAAGAAT CTTTTTGGTA 1920
TCGGGGACAC CCAACTACAT TATTCCCATA TGTTCCCTAT TGCAATGTAT TCTGTGTGCA 1980
ATCTGGCTAG CAGTTTCTCC TCCCTTTGTT GATATTGATG AACACTCTGA GCATGGCCAC 2040
ATCATCATTG TGTGCAACAA GGGCTCCATT ACTGCATTCT ACTGTGTCCT GGGATACTTG 2100
GCCTGCCTGG CCTTTGGAAG CTTCACTATA GCTTTCTTGG CAAAGAACCT GCCTGACACA 2160
TTCAACGAAG CCAAGTTCTT GACCTTCAGC ATGCTAGTGT TCTGCGCTGT CTGGGTCACC 2220

TTCCTCCCTG TCTACCATAG CACCAAGGGC AAGGTCATGG TTGCTGTGGA GATCTTCTCC 2280
ATCTTGGCAT CTAGTGCAGG GATGCTGGGA TGCATCTTTG CACCCAAAGT TTACATCATT 2340
TTAATGAGAC CAGACAGAAA TTCGATCCAC AAAATCAGGG AGAAATCATA TTTC 2394

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2085 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

GTCTACCTGT CTCCACATTT CCTTCAGCTT TCCTATGGAC CTTTCTACTC CATCTTCAGT 60
GATAATGAAC AATATCCTTA TCTCTATCAG ATGGGCCCAA AGGACTCATC ACTAGCATTG 120
GCAATGGTCT CCTTCATAAT TTACTTCAAG TGGAACTGGG TTGGGCTATT TATCTCAGAT 180
GATGATCAAG GCAATCAATT TCTCTCAGAG TTGAAAAAAG AGAGCCAAAC CAAGGATATT 240
TGCTTTGCCT TTGTGAACAT GATATCAGTC AGTGATGTTT CATACTATCA TAAAACTGAA 300
ATGTACTACA ACCAAATTGT GATGTCATCC ACAAAGGTTA TTATCATTTA TGGGGAAACA 360
AACAGTATTA TTGAATTGAG CTTCAGAATG TGGTCATCTC CAGTTAAACA GAGAATATGG 420
GTCACCACAA AACAATTTGA TTGCCCTACC AGTAAGAGAG ACTTAACTCA TGGCACATTC 480
TATGGGACCC TTACATTTCT ACACCACTAT GGTGAGATTT CTGGCTTTAA AAATTTTGTA 540
CAGACACGGT ACAATCTCAG AAGCACAGAT TTATATCTAG TAATGCCAGA GTGGAAATAT 600
TTTAACTATG AAGCCTCAGC ATCTAACTGT AAAATACTGA GAAACTATTT ATCCAATATC 660
TCACTGGAAT GGCTAATGGA ACAGAAATTT GACATGTCAT TTAGTGATTA TAGTCACAAC 720
ATATACAATG CTGTATATGC CATTGCTCAT GCACTCCATG AGAAAGATCT GCAAGAATTT 780
GAAAATCAGG CAATAAACAA TGCGAAAGGA GAAAATACTC ACTGCTTGAA GCTAAACTCA 840
TTTCTGAGAA AGACCCACTT CACTAATTCT CTTGGGAACA GAGTAATTAT GAAACAGAGA 900
GAAGTAGTGC ATGGAGACTA TAATATTGTT CACATGTGGA ATTTCTCACA ACGCCTTGGG 960
ATTAAGGTGA AGATAGGACA ATTCAGCCCA CATTTTCCAC AGGGTCAACA GTTACACTTA 1020
TATGTAGACA TGACTGAGTT GGCTACAGGA AGTAGAAAGA TGCCATCCTC AGTGTGCAGT 1080
GCAGATTGCC ATCCTGGATT CAGAAGAATC TGGAAGGAGG AAATGGCAGC CTGCTGTTTT 1140
GTTTGCAACC CCTGCCCTGA AAATGAAATT TCTAATGAGA CGAATATGGA TCAGTGTGCG 1200
AATTGTCCAG AATACCAGTA TGCCAACACA GAAAAGAACA AATGCATCCA GAAAGGTGTG 1260
ATTGTTCTAA GCTATGAAGA CCCCTTGGGG ATGGCTCTTG CCTTAATAGC ATTCTGTTTC 1320
TCTGCATTCA CAGTGGTGGT ATTTTGGGTC TTCGTGAAGC ACCATGACAC TCCTATTGTG 1380
AAGGCCAATA ACAGAATCCT CAGCTACCTA TTAATCGTGT CACTCATGTT CTGTTTTCTG 1440
TGCTCCTTTT TCTTCATTGG CTATCCTAAC AGAGCAACCT GTATCTTACA GCAAATCACA 1500
TTTGGAATCT TCTTTACTGT GGCTATTTCC ACAGTTCTGG CCAAAACAAT CACTGTGGTT 1560
CTGGCTTTCA AAGTCACAGA CCCAGGAAGA CAATTAAGAA TCTTTTTGGT ATCGGGGACA 1620
CCCAACTACA TTATTCCCAT ATGTTCCCTA TTGCAATGTA TTCTGTGTGC AATCTGGCTA 1680
GCAGTTTCTC CTCCCTTTGT TGATATTGAT GAACACTCTG AGCATGGCCA CATCATCATT 1740
GTGTGCAACA AGGGCTCCAT TACTGCATTC TACTGTGTCC TGGGATACTT GGCCTGCCTG 1800
GCCTTTGGAA GCTTCACTAT AGCTTTCTTG GCAAAGAACC TGCCTGACAC ATTCAACGAA 1860
GCCAAGTTCT TGACCTTCAG CATGCTAGTG TTCTGCGCTG TCTGGGTCAC CTTCCTCCCT 1920
GTCTACCATA GCACCAAGGG CAAGGTCATG GTTGCTGTGG AGATCTTCTC CATCTTGGCA 1980
TCTAGTGCAG GGATGCTGGG ATGCATCTTT GCACCCAAAG TTTACATCAT TTTAATGAGA 2040
CCAGACAGAA ATTCGATCCA CAAAATCAGG GAGAAATCAT ATTTC 2085

We claim:

Claims

1. A family of pheromone receptor polypeptides, each of said polypeptides comprising from amino terminus to carboxyl terminus:

(a) an amino-terminal extracellular domain containing from 30 to 600 amino acids;

(b) a transmembrane region comprising:

(i) seven non-contiguous transmembrane domains designated TM1, TM2, TM3, TM4, TM5, TM6 and TM7

(ii) three non-contiguous extracellular domains designated EC2, EC3 and EC4, and

(iii) three non-contiguous intracellular domains designated IC1, IC2, and IC3,

wherein the transmembrane domains, the extracellular domains and the intracellular domains are attached to one another from amino terminus to carboxyl terminus in the order TM1-IC1-TM2-EC2-TM3- IC2-TM4-EC3-TM5-IC3-TM6-EC4-TM7, and

wherein the transmembrane region has at least about 35% homology and a length approximately equal to a transmembrane region of a polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 34, 36, 38, 40, 42, 44, 46, 48, and 50; and

(c) a carboxyl-terminal intracellular domain containing from 5 to 200 amino acids;

wherein the pheromone receptor polypeptides are expressed in a $G\alpha_o$ protein-expressing vomeronasal organ neuron or are expressed in another olfactory organ neuron in an animal which does not possess a vomeronasal organ.

2. The polypeptides of claim 1, wherein the transmembrane region of each of said polypeptides has at least between about 60% and about 90% homology to the transdomain region of a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 34, 36, 38, 40, 42, 44, 46, 48, and 50.

3. The polypeptides of claims 1 or 2 , wherein the non-contiguous intracellular domains of each of said polypeptides has at least between about 60% and about 90% homology to the non-contiguous intracellular domains of a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 34, 36, 38, 40, 42, 44, 46, 48, and 50.

4. The polypeptides of claim 1, wherein the extracellular domain of each of said polypeptides has at least between about 50% and about 90% homology to the extracellular domain of a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, and 50.

5. The polypeptides of claim 2, wherein the extracellular domain of each of said polypeptides has at least between about 50% and about 90% homology to the extracellular domain of a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, and 50.

6. The polypeptides of claim 3, wherein the extracellular domain of each of said polypeptides has at least between about 50% and about 90% homology to the extracellular domain of a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, and 50 .

7. The polypeptides of claims 1 or 2, wherein the extracellular domain contains at least between about 50 and about 500 amino acids.

8. The polypeptides of claim 3, wherein the extracellular domain contains at least between about 50 and about 500 amino acids.

9. The polypeptides of claims 4, 5 or 6, further comprising a signal sequence attached to the amino terminus of the extracellular domain.

10. The polypeptides of claim 9, wherein the signal sequence is selected from the group of signal sequences of a pheromone receptor polypeptide of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52.

11. A method for identifying a nucleic acid encoding a pheromone receptor polypeptide, comprising:

(1) contacting a mixture of nucleic acid molecules with at least one nucleic acid probe of a nucleic acid selected from the group consisting of: (a) a nucleic acid molecule selected from

the group consisting of SEQ ID NO. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 54, and 55 that encodes a pheromone receptor polypeptide; (b) a unique fragment of (a); (c) a human homolog of (a) or (b); and (d) a set of degenerate primers of any of (a), (b) or (c); and

(2) identifying the sequences within the mixture that hybridize to the probe.

12. The method of claim 11, wherein the mixture is a genomic library.

13. The method of claim 11, wherein the mixture is a cDNA library.

14. The method of claim 11, wherein the nucleic acid probe contains a detectable label.

15. The method of claim 11, wherein the at least one nucleic acid probe is a pair of degenerate polymerase chain reaction primers that amplify a unique fragment of a nucleic acid molecule selected from the group consisting of SEQ ID NO. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 54, and 55, the method further comprising the step of subjecting the mixture to a polymerase chain reaction amplification reaction prior to selecting a member of the mixture which hybridizes to the nucleic acid probe.

16. The method of claim 15, wherein the pair of degenerate polymerase chain reaction primers is selected from the group consisting of SEQ ID NOs. 60 and 61, SEQ ID NOs. 62 and 63, SEQ ID NOs. 64 and 63, SEQ ID NOs. 64 and 65, and SEQ ID NOs. 66 and 67.

17. The method of claim 16, wherein the pair of polymerase chain reaction primers is selected from the group consisting of SEQ ID NOs. 60 and 61, SEQ ID NOs. 62 and 63, SEQ ID and NOs. 64 and 63.

18. An isolated nucleic acid molecule

(a) which hybridizes under high or low stringency conditions to a molecule consisting of a nucleic acid sequence selected from the group consisting of SEQ ID NO. 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 54, and 55, and which codes for a pheromone receptor,

(b) nucleic acid molecules that differ from the nucleic acid molecules of (a) in codon sequence due to the degeneracy of the genetic code, and

(c) complements of (a) and (b).

19. The nucleic acid molecule of claim 18, wherein the pheromone receptor is expressed in the vomeronasal organ or is expressed in another olfactory organ in an animal which does not possess a vomeronasal organ.

20. The nucleic acid molecule of claim 18, wherein the pheromone receptor is expressed in a $G\alpha_o$ protein-expressing vomeronasal organ neuron.

21. The nucleic acid molecule of claim 18, wherein the pheromone receptor is a G-protein coupled receptor.

22. The isolated nucleic acid molecule of claim 18, wherein the pheromone receptor has an amino acid sequence selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52.

23. The isolated nucleic acid molecule of claim 18, wherein the isolated nucleic acid molecule is selected from the group consisting of SEQ ID NO. 51, 53, 54, 55, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, and 92, that encodes a pheromone receptor polypeptide.

24. The isolated nucleic acid molecule of claim 18, wherein the isolated molecule comprises a molecule having a sequence which encodes a pheromone receptor unique fragment, wherein said unique fragment is selected from the group consisting of a pheromone receptor extracellular domain, a pheromone receptor transmembrane domain, a pheromone receptor intracellular domain, a pheromone receptor extracellular domain coupled to at least one transmembrane domain, and at least one pheromone receptor transmembrane domain coupled to a pheromone receptor intracellular domain.

25. The isolated nucleic acid molecule of claim 18, wherein the pheromone receptor extracellular domain, the pheromone receptor transmembrane domain and the pheromone receptor intracellular domain have amino acid sequences selected from the group of sequences identified as these domains in SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52.

26. The isolated nucleic acid molecule of claim 18, wherein the unique fragment is selected from the group consisting of between 12 and 4000, between 12 and 2000, between 12 and 1000, between 12 and 500, between 12 and 250, between 12 and 100, between 12 and 50, and between 12 and 25, nucleotides in length.

27. An isolated nucleic acid molecule, comprising

(a) a molecule having a sequence selected from the group consisting of SEQ ID NO. 51, 53, 54, 55, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, and 92, and which codes for a pheromone receptor;

(b) nucleic acid molecules that differ from the nucleic acid molecules of (a) in codon sequence due to the degeneracy of the genetic code, and

(c) complements of (a) and (b).

28. An expression vector comprising the isolated nucleic acid molecule of claims 18-27 operably linked to a promoter.

29. A host cell transformed or transfected with the isolated nucleic acid molecule of claims 18-27.

30. A host cell transformed or transfected with the isolated nucleic acid molecule of the expression vector of claim 28.

31. An isolated polypeptide encoded by the isolated nucleic acid molecule of claims 18-27.

32. The isolated polypeptide of claim 31, wherein the isolated polypeptide has a pheromone receptor activity.

33. The isolated polypeptide of claim 31, wherein the isolated polypeptide comprises a polypeptide selected from group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52.

34. The isolated polypeptide of claim 33, wherein the isolated polypeptide is a fragment of a peptide selected from the group consisting of an extracellular domain, a transmembrane domain and an intracellular domain, wherein the foregoing domains have amino acid sequences selected from the group of sequences identified as these domains of a pheromone receptor polypeptide selected from group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52.

35. A vaccine containing an isolated polypeptide selected from the group consisting of the isolated polypeptides of claim 31, 32, 33, and 34.

36. A method for controlling fertility in an animal, comprising:

administering to an animal in need of such treatment, an effective amount of the vaccine of claim 35 to elicit an immune response to the isolated polypeptide.

37. An isolated binding polypeptide which binds selectively to a polypeptide of claim 1, 2, 4, 5, 6, 8, 10, 31, 32, 33, and 34, provided that the isolated binding polypeptide does not bind to a G-protein coupled receptor other than a $G\alpha_0^+$-coupled pheromone receptor.

38. The isolated binding polypeptide of claim 37, wherein the binding polypeptide binds to a polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50 and 52.

39. The isolated binding polypeptide of claim 37, wherein the binding polypeptide is an antibody fragment selected from the group consisting of a Fab fragment, a $F(ab)_2$ fragment or a fragment including a CDR3 region selective for *a pheromone receptor* polypeptide.

40. The isolated binding polypeptide of claim 38, wherein the binding polypeptide is an antibody fragment selected from the group consisting of a Fab fragment, a $F(ab)_2$ fragment or a fragment including a CDR3 region selective for *a pheromone receptor* polypeptide.

41. An affinity matrix comprising:

a solid support to which is coupled an isolated binding polypeptide selected from the group consisting of the binding polypeptides of any of claims 37-40.

42. A method for isolating a pheromone receptor, comprising:

contacting a composition containing a putative pheromone receptor with the affinity matrix of claim 41 under conditions to permit the pheromone receptor to selectively bind to the binding polypeptides coupled to the solid support; and

isolating the polypeptides that bind to the affinity matrix.

43. A composition comprising:

the polypeptide of claim 1, 2, 4, 5, 6, 8, 10, 31, 32, 33, or 34; and

a pharmaceutically acceptable carrier.

44. A composition comprising:

the nucleic acid molecule of any of claims 18-28; and

a pharmaceutically acceptable carrier.

45. A composition comprising:

the binding polypeptide of claim 37; and

a pharmaceutically acceptable carrier.

46. A composition comprising:

the binding polypeptide of claims 38, 39 or 40; and

a pharmaceutically acceptable carrier.

47. A method for modulating a pheromone receptor activity in a cell, comprising:

administering to the cell an amount of the isolated binding polypeptide of claim 37 effective to modulate pheromone receptor activity in the cell.

48. A method for modulating a pheromone receptor activity in a cell, comprising:

administering to the cell an amount of the isolated binding polypeptide of claim 38, 39, or 40 effective to modulate pheromone receptor activity in the cell.

49. The method of claim 47, wherein modulating a pheromone receptor activity comprises reducing the pheromone receptor activity.

50. The method of claim 48, wherein modulating a pheromone receptor activity comprises reducing the pheromone receptor activity.

51. The method of claim 47, wherein the pheromone receptor activity is selected from the group consisting of a signal transduction activity and a ligand binding activity.

52. The method of claim 48, wherein the pheromone receptor activity is selected from the group consisting of a signal transduction activity and a ligand binding activity.

53. The method of claim 47, wherein the cell is a vertebrate cell, preferably a mammalian cell.

54. The method of claim 48, wherein the cell is a vertebrate cell, preferably a mammalian cell.

55. The method of claim 47, wherein the cell is an invertebrate cell, preferably an insect cell.

56. The method of claim 48, wherein the cell is an invertebrate cell, preferably an insect cell.

57. A method for reducing the binding of a pheromone having a binding domain to a pheromone receptor having a ligand binding site that selectively binds to the binding domain of the pheromone, comprising:

contacting the pheromone receptor with an agent which binds to the binding domain for a time effective to reduce binding of the pheromone to the ligand binding site of the pheromone receptor.

58. The method of claim 57, wherein the agent is an antibody which binds to the binding domain.

59. A method for decreasing pheromone receptor mediated signal transduction activity in a subject comprising:

administering to a subject in need of such treatment an agent that selectively binds to an isolated nucleic acid molecule of claim 1 or an expression product thereof, in an amount effective to decrease pheromone receptor mediated signal transduction activity in the subject.

60. The method of claim 59, wherein the agent is selected from the group consisting of an antisense nucleic acid and a binding polypeptide.

61. A method for identifying lead compounds for a pharmacological agent useful in the diagnosis or treatment of disease associated with pheromone binding to a pheromone receptor polypeptide containing a ligand binding site that selectively binds to a binding domain of the pheromone, comprising

forming a mixture comprising a pheromone receptor polypeptide or unique fragment thereof containing a ligand binding site, a molecule protein containing a binding domain which selectively binds the pheromone receptor ligand binding site, and a candidate pharmacological agent,

incubating the mixture under conditions which, in the absence of the candidate pharmacological agent, permit a first amount of selective binding of the molecule containing a ligand binding domain by the pheromone receptor ligand binding site, and

detecting a test amount of selective binding of the molecule containing the binding domain by the pheromone receptor ligand binding site, wherein reduction of the test amount of selective binding relative to the first amount of selective binding indicates that the candidate pharmacological agent is a lead compound for a pharmacological agent which disrupts selective

binding of a molecule containing a binding domain by a pheromone receptor containing a ligand binding site and wherein increase of the test amount of selective binding relative to the first amount of selective binding indicates that the candidate pharmacological agent is a lead compound for a pharmacological agent which enhances selective binding of a molecule containing a binding domain by a pheromone receptor polypeptide containing a ligand binding site.

AMENDED CLAIMS

[received by the International Bureau on 11 December 1998 (11.12.98); original claim 1 amended; remaining claims unchanged (1 page)]

1. A family of isolated pheromone receptor polypeptides, each of said isolated polypeptides comprising from amino terminus to carboxyl terminus:

(a) an amino-terminal extracellular domain containing from 30 to 600 amino acids;

(b) a transmembrane region comprising:

(i) seven non-contiguous transmembrane domains designated TM1, TM2, TM3, TM4, TM5, TM6 and TM7

(ii) three non-contiguous extracellular domains designated EC2, EC3 and EC4, and

(iii) three non-contiguous intracellular domains designated IC1, IC2, and IC3,

wherein the transmembrane domains, the extracellular domains and the intracellular domains are attached to one another from amino terminus to carboxyl terminus in the order TM1-IC1-TM2-EC2-TM3- IC2-TM4-EC3-TM5-IC3-TM6-EC4-TM7, and

wherein the transmembrane region has at least about 35% homology and a length approximately equal to a transmembrane region of a polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 34, 36, 38, 40, 42, 44, 46, 48, and 50; and

(c) a carboxyl-terminal intracellular domain containing from 5 to 200 amino acids;

wherein the pheromone receptor polypeptides are expressed in a $G\alpha_o$ protein-expressing vomeronasal organ neuron or are expressed in another olfactory organ neuron in an animal which does not possess a vomeronasal organ.

2. The polypeptides of claim 1, wherein the transmembrane region of each of said polypeptides has at least between about 60% and about 90% homology to the transdomain region of a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 12, 14, 34, 36, 38, 40, 42, 44, 46, 48, and 50.

3. The polypeptides of claims 1 or 2 , wherein the non-contiguous intracellular domains of each of said polypeptides has at least between about 60% and about 90% homology to the non-contiguous intracellular domains of a pheromone receptor polypeptide selected from the group consisting of SEQ ID NO. 2, 4, 6, 8, 10, 34, 36, 38, 40, 42, 44, 46, 48, and 50.

FIGURE 1.

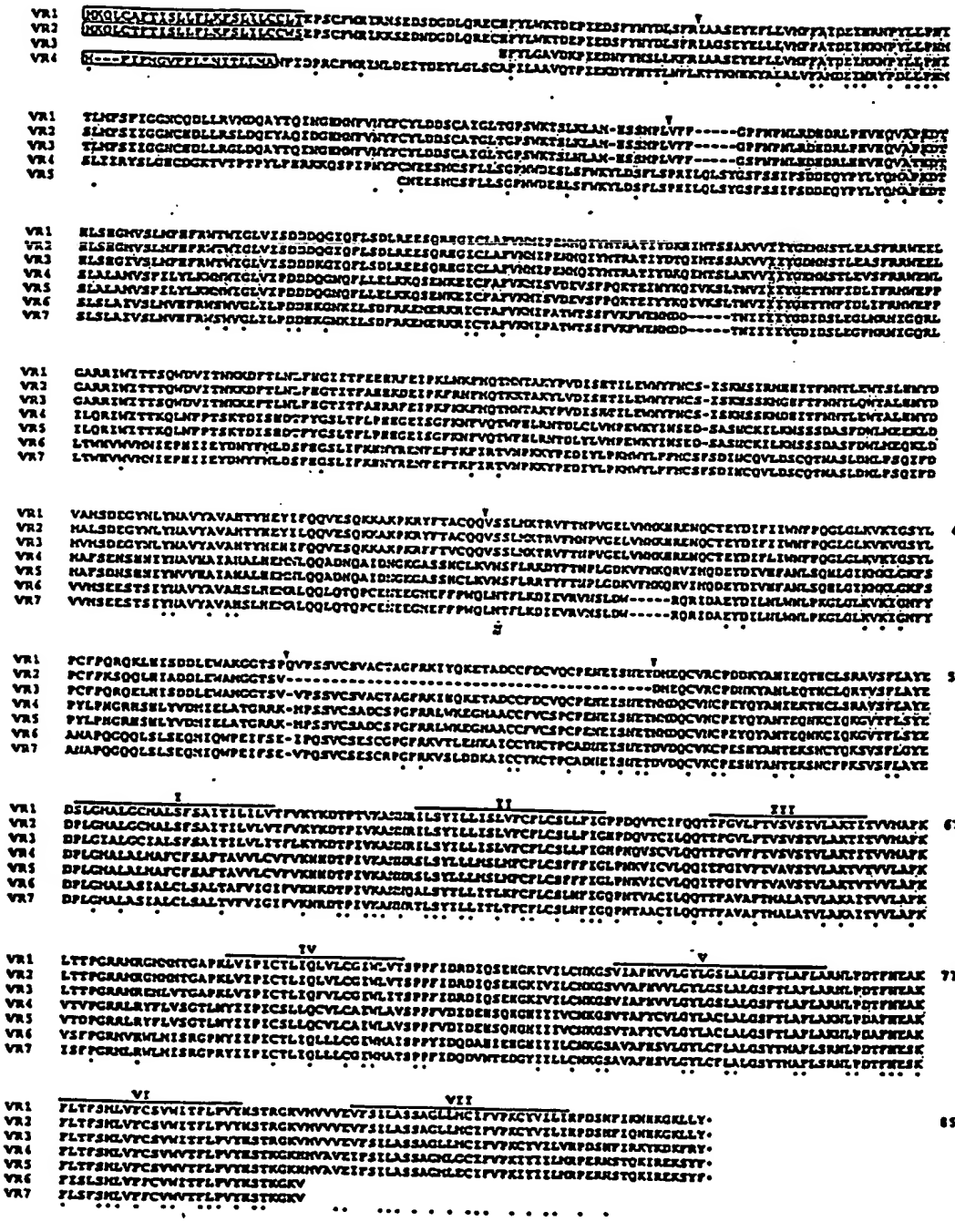
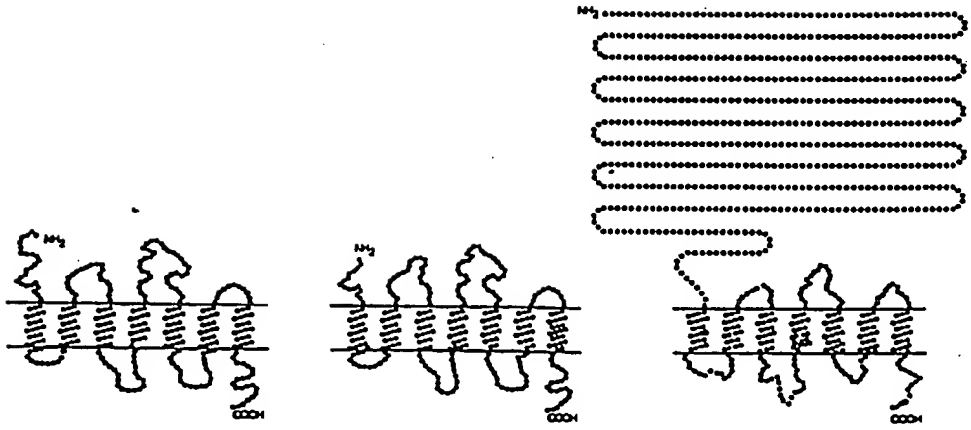


FIGURE 2.



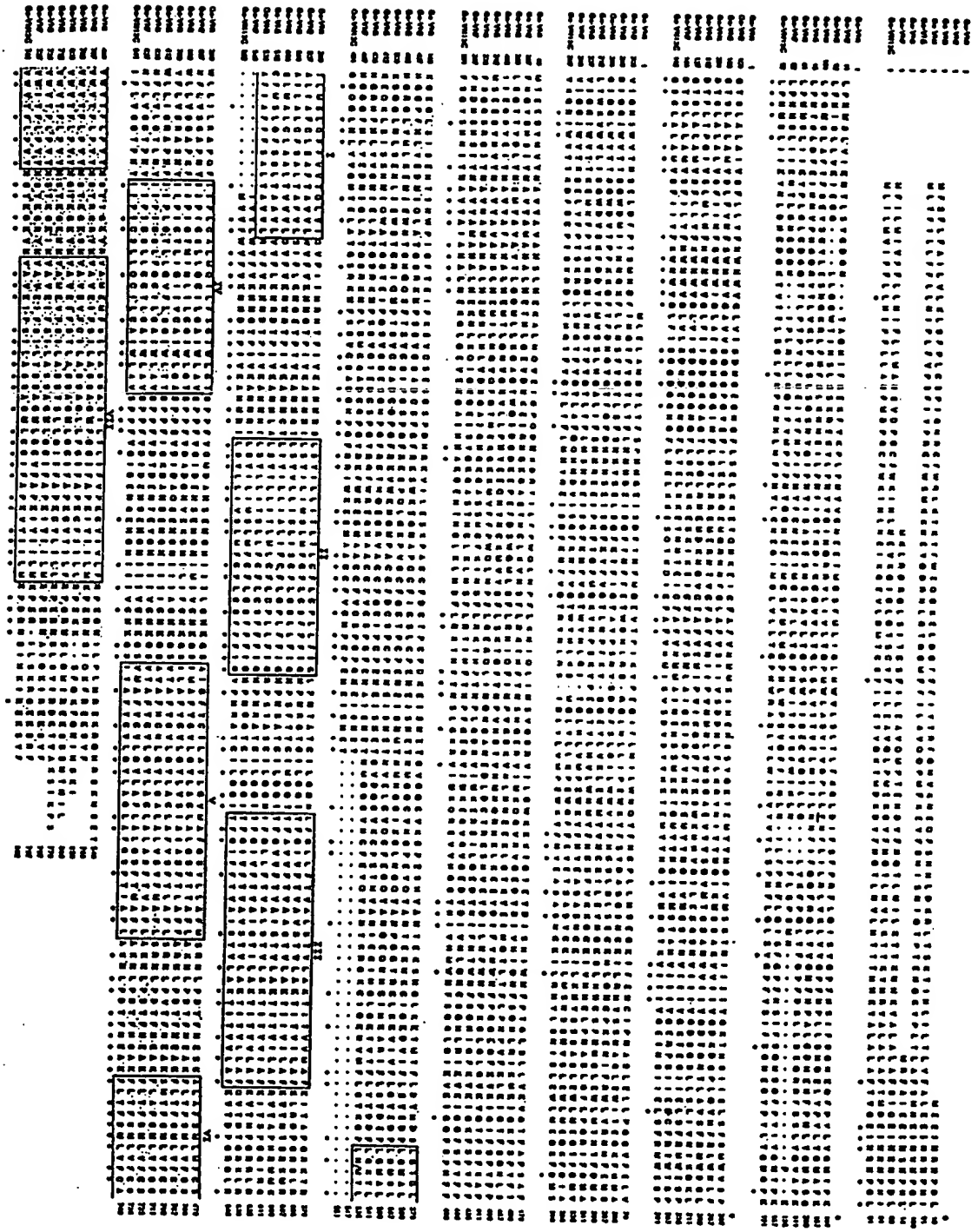


FIGURE 3.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/13680

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C07K 14/705; C12N 15/12; A61K 38/17; C12Q 1/68
US CL :536/23.5, 24.31; 530/350; 514/2; 435/6
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.5, 24.31; 530/350; 514/2; 435/6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Biosis, Medline, WPI
search terms: pheromone receptor, odorant receptor, vomeronasal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	BROWN et al. Cloning and Characterization of an Extracellular Ca^{2+}-Sensing Receptor from Bovine Parathyroid. Nature. 09 December 1993, Vol. 366, pages 575-580, pages 577 and 578.	18-21, 24, 26 ---------- 1-17, 22, 23, 25, 27, 43
A	KIEFER et al. Expression of an Olfactory Receptor in Escherichia coli: Purification, Reconstitution, and Ligand Binding. Biochemistry. 1996, Vol. 35, No. 50, pages 16077-16084.	1-27, 43

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"B"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 18 SEPTEMBER 1998	Date of mailing of the international search report OCT 13 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer W. Laurence for SALLY P. TENG Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/13680

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, P	HERRADA et al. A Novel Family of Putative Pheromone Receptors in Mammals with a Topographically Organized and Sexually Dimorphic Distribution. Cell. 22 August 1997, Vol. 90, pages 763-773, see pages 765-767.	1-27, 43 (Species 17)
X, P	MATSUNAMI et al. A Multigene Family Encoding a Diverse Array of Putative Pheromone Receptors in Mammals. Cell. 22 August 1997, Vol. 90, pages 775-784, pages 776-778.	1-27, 43 (species 1 and 4)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/13680

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 28-42, 44-56
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
1-27 and 43, species 1, 4, 17, 26-29

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/13680

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claims 1-27, 43, drawn to pherome receptor polypeptides and their encoding nucleic acids.
Group II, claims 57 and 58, drawn to a method of reducing the binding of a pheromone to a pherome receptor.
Group III, claims 59 and 60, drawn to a method of decreasing pherome receptor mediated signal transduction.
Group IV, claim 61, drawn to a method of identifying lead compounds.

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack Unity of Invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for more than one species to be searched, the appropriate additional search fees must be paid. The species are as follows:

1) SEQ ID NO: 1 and 2;
2) SEQ ID NO: 3 and 4;
3) SEQ ID NO: 5 and 6;
4) SEQ ID NO: 7 and 8;
5) SEQ ID NO: 9 and 10;
6) SEQ ID NO: 11 and 12;
7) SEQ ID NO: 13 and 14;
8) SEQ ID NO: 15 and 16;
9) SEQ ID NO: 17 and 18;
10) SEQ ID NO: 19 and 20;
11) SEQ ID NO: 21 and 22;
12) SEQ ID NO: 23 and 24;
13) SEQ ID NO: 25 and 26;
14) SEQ ID NO: 27 and 28;
15) SEQ ID NO: 29 and 30;
16) SEQ ID NO: 31 and 32;
17 SEQ ID NO: 33 and 34;
18 SEQ ID NO: 35 and 36;
19) SEQ ID NO: 37 and 38;
20) SEQ ID NO: 39 and 40;
21) SEQ ID NO: 41 and 42;
22) SEQ ID NO: 43 and 44;
23) SEQ ID NO: 45 and 46;
24) SEQ ID NO: 47 and 48;
25) SEQ ID NO: 49 and 50;
26) SEQ ID NO: 51 and 52;
27) SEQ ID NO: 53;
28) SEQ ID NO: 54;
29) SEQ ID NO: 55;
30) SEQ ID NO: 68;
31) SEQ ID NO: 69;
32) SEQ ID NO: 70;
33) SEQ ID NO: 71;
34) SEQ ID NO: 72;
35) SEQ ID NO: 73;
36) SEQ ID NO: 74;
37) SEQ ID NO: 75;
38) SEQ ID NO: 76;
39) SEQ ID NO: 77;
40) SEQ ID NO: 78;
41) SEQ ID NO: 79;
42) SEQ ID NO: 80;
43) SEQ ID NO: 81;
44) SEQ ID NO: 82;
45) SEQ ID NO: 83;

46) SEQ ID NO: 84;
47) SEQ ID NO: 85;
48) SEQ ID NO: 86;
49) SEQ ID NO: 87;
50) SEQ ID NO: 88;
51) SEQ ID NO: 89;
52) SEQ ID NO: 90;
53) SEQ ID NO: 91;
54) SEQ ID NO: 92.

The claims are deemed to correspond to the species listed above in the following manner:

The claims are directed to pheromone receptor polypeptides and their encoding nucleic acids having the recited sequences.

The following claims are generic: 1-27, 43, and 57-61.

The inventions listed as Groups I-IV do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: It is noted that the expression "special technical features" is defined in Rule 13.2 as meaning "those technical features that define a contribution which each of the inventions, considered as a whole makes over the prior art". The claimed invention of Group I, directed to a family of pheromone receptor polypeptide, encompasses naturally occurring non-isolated products present in the vomeronasal organ and is anticipated by the prior art (see Dulac and Axel). Therefore, the polypeptide of Group I lacks a special technical feature. The special technical feature of Group II is a method of using a binding protein to reduce the binding of the pheromone receptor to its ligand. The special technical feature of Group III is a method of using a compound that binds to the nucleic acid encoding a pheromone receptor to decrease pheromone receptor mediated signal transduction. The special technical feature of Group IV is a method of identifying lead compounds for a pharmacological agent useful in the diagnosis or treatment of disease associated with pheromone binding to a pheromone receptor. The special technical feature of each group is not the same or does not correspond to the special technical feature of any other group because the methods of Groups II, III, and IV require different starting reagents and method steps to accomplish different goals. The Groups are not linked by a special technical feature within the meaning of PCT Rule 13.2 so as to form a single inventive concept.

The species listed above do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: Each of the species has a distinct amino acid sequence and is encoded by a distinct nucleic acid sequence.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/13680

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 28-42, 44-56
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
1-27 and 43, species 1, 4, 17, 26-29

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.